
Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
10.1136/bmj.j4761

Link to publication record in Explore Bristol Research

PDF-document

This is the final published version of the article (version of record). It first appeared online via BMJ at http://www.bmj.com/content/359/bmj.j4761. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms
Circulating vitamin D concentration and risk of seven cancers: Mendelian randomisation study

Vasiliki I Dimitrakopoulou,1,2 Konstantinos K Tsilidis,1,3 Philip C Haycock,4,5 Niki L Dimou,1 Kawthar Al-Dabhani,3 Richard M Martin,4,6 Sarah J Lewis,4,5 Marc J Gunter,7 Alison Mondul,8 Irene M Shui,9 Evropi Theodoratou,10 Katharina Nimptsch,11 Sara Lindström,12 Demetrius Albanes,13 Tilman Kühn,14 Timothy J Key,15 Ruth C Travis,15 Karani Santhanakrishnan Vimaleswaran,16 Peter Kraft,17 Brandon L Pierce,18,20 Joellen M Schildkraut21

For numbered affiliations see end of article.
Correspondence to: K K Tsilidis ktsilidis@cc.uoi.gr
Additional material is published online only. To view please visit the journal online.
Cite this as: BMJ 2017;359:j4761 http://dx.doi.org/10.1136/bmj.j4761
Accepted: 26 September 2017

ABSTRACT

OBJECTIVE
To determine if circulating concentrations of vitamin D are causally associated with risk of cancer.

DESIGN
Mendelian randomisation study.

SETTING
Large genetic epidemiology networks (the Genetic Associations and Mechanisms in Oncology (GAME-ON), the Genetic and Epidemiology of Colorectal Cancer Consortium (GECCO), and the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortiums, and the MR-Base platform).

PARTICIPANTS
70,563 cases of cancer (22,898 prostate cancer, 15,748 breast cancer, 12,537 lung cancer, 11,488 colorectal cancer, 4,369 ovarian cancer, 1,896 pancreatic cancer, and 1,627 neuroblastoma), and 84,418 controls.

EXPOSURES
Four single nucleotide polymorphisms (rs2282679, rs10741657, rs12785878 and rs6013897) associated with vitamin D were used to define a multi-polymorphism score for circulating 25-hydroxyvitamin D (25(OH)D) concentrations.

MAIN OUTCOMES MEASURES
The primary outcomes were the risk of incident colorectal, breast, prostate, ovarian, lung, and pancreatic cancer and neuroblastoma, which was evaluated with an inverse variance weighted average of the associations with specific polymorphisms and a likelihood based approach. Secondary outcomes based on cancer subtypes by sex, anatomic location, stage, and histology were also examined.

RESULTS
There was little evidence that the multi-polymorphism score of 25(OH)D was associated with risk of any of the seven cancers or their subtypes. Specifically, the odds ratios per 25 nmol/L increase in genetically determined 25(OH)D concentrations were 0.92 (95% confidence interval 0.77 to 1.08) for colorectal cancer, 1.05 (0.89 to 1.24) for breast cancer, 0.89 (0.77 to 1.02) for prostate cancer, and 1.03 (0.87 to 1.23) for lung cancer. The results were consistent with the two different analytical approaches, and the study was powered to detect relative effect sizes of moderate magnitude (for example, 1.20-1.50 per 25 nmol/L decrease in 25(OH)D for most primary cancer outcomes. The Mendelian randomisation assumptions did not seem to be violated.

CONCLUSIONS
There is little evidence for a linear causal association between circulating vitamin D concentration and risk of various types of cancer, though the existence of causal clinically relevant effects of low magnitude cannot be ruled out. These results, in combination with previous literature, provide evidence that population-wide screening for vitamin D deficiency and subsequent widespread vitamin D supplementation should not currently be recommended as a strategy for primary cancer prevention.

WHAT IS ALREADY KNOWN ON THIS TOPIC
There is debate about whether vitamin D status is linked with disease or is just a correlate marker of overall health
Evidence from in vitro and animal model studies supports an anti-neoplastic role of vitamin D, but epidemiological studies and randomised controlled trials have yielded mixed results

WHAT THIS STUDY ADDS
This Mendelian randomisation study provides little evidence of a linear causal association between circulating vitamin D concentration and risk of colorectal, breast, prostate, ovarian, lung, and pancreatic cancer and neuroblastoma, but the existence of causal clinically relevant effects of low magnitude cannot be ruled out
Population-wide screening for vitamin D deficiency and subsequent widespread vitamin D supplementation cannot currently be recommended as a strategy for primary cancer prevention

Introduction
Evidence from in vitro and animal model studies supports an anti-neoplastic role of vitamin D.1 Vitamin D functions by activating the nuclear vitamin D receptor, which is ubiquitously expressed and regulates the growth, differentiation, and apoptosis of normal and tumour cells.1 Epidemiological studies of circulating vitamin D concentrations and risk of various cancers have produced inconsistent results. Meta-analyses of observational studies have suggested that higher concentrations of 25-hydroxyvitamin D (25(OH)D), the primary circulating form, is associated with a lower risk of colorectal cancer.2 Epidemiological evidence for breast and prostate cancer is inconclusive, while data for other cancers are limited.3,4 Previous observational associations between circulating 25(OH)D and cancer are limited by relatively small study specific sample
scales (for example, 3000-5000 cases in meta-analyses of breast, prostate, and colorectal cancer) and by several potential methodological issues. Specifically, reverse causation could exist if 25(OH)D is measured at or close to cancer diagnosis, residual confounding might be present because of inadequate control for common causes of cancer, and errors in measurement of exposure to 25(OH)D could result from single measurements.

Definitive data from randomised controlled trials are lacking as few adequately powered trials have examined vitamin D supplementation and risk of cancer. The Women's Health Initiative, a randomised placebo controlled trial of 400 IU of vitamin D plus 1000 mg of calcium per day in 36 282 postmenopausal women, failed to support a protective role of vitamin D over a period of seven years for colorectal cancer (n=332 cases), breast cancer (n=1074 cases), or total cancer (n=2639 cases). The dose of vitamin D, however, was probably inadequate and the follow-up was too short to yield a substantial contrast. A meta-analysis of four vitamin D supplementation trials found no association with total cancer incidence. Another meta-analysis of 18 trials found a decrease in total cancer mortality, but the possibility of type I error and attrition bias was reported as few participants were examined and there was substantial dropout.

A previous Mendelian randomisation study reported that genetically low 25(OH)D concentrations were associated with increased cancer mortality, but this study included only about 2800 deaths from cancer and could not perform analyses by cancer site.

To overcome limitations of conventional observational research and randomised trials and shed light on whether vitamin D status is a cause of disease or just a correlate marker of overall health, we used a Mendelian randomisation approach and estimated associations between single nucleotide polymorphisms associated with vitamin D and risk of colorectal, breast, prostate, ovarian, lung, and pancreatic cancer and neuroblastoma using summary data from the Genetic Associations and Mechanisms in Oncology (GAME-ON), the Genetic and Epidemiology of Colorectal Cancer Consortium (GECCO), and the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortiums, and the MR-Base platform. Mendelian randomisation aims to improve causal inference by assessing risk associations of the genetically determined component of environmental exposures and biomarkers.

**Methods**

**Data for genetic epidemiology of cancer**

We retrieved summary data for the association between single nucleotide polymorphisms associated with vitamin D and cancer from three large genetic epidemiology networks. The GAME-ON initiative is a network of five cancer specific consortiums: CORECT (ColoRectal Transdisciplinary Study); DRIVE (Discovery, Biology, and Risk of Inherited Variants in Breast Cancer); ELLIPSE (Elucidating Loci Involved in Prostate Cancer Susceptibility); FOCI-OCAC (Follow-up of Ovarian Cancer Genetic Association and Interaction Studies of the Ovarian Cancer Association Consortium); and TRICL-ILCCO (Transdisciplinary Research in Cancer of the Lung of the International Lung Cancer Consortium). Larger scale summary data on the genetic epidemiology of colorectal and prostate cancer were retrieved from the GECCO and PRACTICAL consortiums. Further details on these networks can be found elsewhere. Data for the genetic epidemiology of pancreatic cancer and neuroblastoma were retrieved from PanScan1 (Pancreatic Cancer Cohort Consortium Genome-Wide Association Study) and from a genome-wide association study of neuroblastoma through the MR-Base platform.

Results from individual genome-wide association studies for each cancer type were combined by using standard fixed effects meta-analysis methods. We used Illumina or Affymetrix arrays for genotyping and either MACH or IMPUTE for imputation with the 1000 Genomes reference panel. We incorporated principal components as covariates in the single nucleotide polymorphism and cancer logistic regression models to adjust for population stratification. Further information regarding the statistical analysis, imputation, and quality control steps in the genome-wide association studies have been previously reported.

**Data for genetic epidemiology of circulating 25(OH) D concentrations**

We conducted a search of published genome-wide association studies in PubMed and the relevant catalogue and identified four single nucleotide polymorphisms as robustly associated at $P<5 \times 10^{-8}$ with circulating 25(OH)D concentration in two genome-wide association studies. These were rs2282679 in the group specific component (GC) on chromosome 4p12 that encodes the vitamin D binding protein; rs10741657 in CYP2R1 on chromosome 11p15 that is involved in the hydroxylation of vitamin D3 to 25(OH)D; rs12785878 located near DHCR7 on chromosome 11q12 that catalyses the conversion of 7-dehydrocholesterol to cholesterol; and rs6013897 near CYP24A1 on chromosome 20q13 that encodes an enzyme that initiates the degradation of 1,25(OH)D. All four single nucleotide polymorphisms were identified among individuals of European ancestry. Each explained about 1% of the 25(OH)D variability, and up to 5% for combinations of the four polymorphisms. More precise estimates of the associations between the polymorphisms with circulating 25(OH)D were obtained from a recent large Mendelian randomisation study. That study analysed the association between the aforementioned polymorphisms with circulating 25(OH)D concentrations as per unit change (nmol/L) in the natural (untransformed) scale.

**Statistical analysis**

We conducted Mendelian randomisation analyses to test the potential causal associations between circulating 25(OH)D and the risk of seven cancers...
(colorectal, breast, prostate, ovarian, lung, pancreatic, and neuroblastoma) using summary data from GAME-ON, GECCO, PRACTICAL, MR-Base, and genome-wide association studies of 25(OH)D concentration. We also performed analyses for cancer subtypes: colorectal cancer in men and women, colon cancer, rectal cancer, proximal colon cancer, distal colon cancer, oestrogen receptor negative breast cancer; aggressive prostate cancer; clear cell, endometrioid, and serous ovarian cancer; and adenocarcinoma and squamous cell carcinoma of lung. We formulated a weighted multi-polymorphism score, which has been previously shown to be linearly associated with circulating 25(OH)D concentration. We used two Mendelian randomisation methods using summary genetic data: an inverse variance weighted average of associations for specific polymorphisms and a likelihood based method. More information on these methods is provided in appendix 1.

For the Mendelian randomisation analyses to have a valid interpretation, it is necessary that the following three instrumental variable assumptions hold: the genetic markers are strongly associated with circulating vitamin D concentration; the markers affect cancer only through their effect on circulating vitamin D; and markers are independent of any confounders of the association between circulating vitamin D and cancer. To assess potential violation of these assumptions we performed several statistical tests (MR-Egger, weighted median approach, and over-identification tests) and sensitivity analyses (Mendelian randomisation analyses with two separate allelic scores: a vitamin D synthesis allele score (rs10741657 and rs12785878) and a metabolism allele score (rs2282679 and rs6013897)), more information about which is provided in appendix 1.

**Results**

**Descriptives and statistical power**

Table 1 lists the sample sizes used in the current study for each cancer type. The number of cancer cases ranged from 1627 for neuroblastoma to 22898 for prostate cancer. Our Mendelian randomisation analyses had 80% power, assuming that 3% of the 25(OH)D variance was explained by the four single nucleotide polymorphisms, to detect effect sizes of moderate magnitude, ranging from odds ratios of 0.58 per SD (for instance, 25 nmol/L or 10 ng/mL) increase in circulating 25(OH)D concentration for neuroblastoma to 0.86 for prostate cancer (table 1), which are comparable with effect sizes that have been observed in observational studies relating circulating

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Study</th>
<th>Cases</th>
<th>Controls</th>
<th>Minimum detectable OR* (R²=0.03)</th>
<th>Minimum detectable OR* (R²=0.05)</th>
<th>OR (95% CI) in published meta-analyses†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal</td>
<td>All</td>
<td>GAME-ON</td>
<td>5100</td>
<td>4831</td>
<td>0.73/1.39</td>
<td>0.78/1.28</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>GECCO</td>
<td>11488</td>
<td>11679</td>
<td>0.81/1.23</td>
<td>0.85/1.18</td>
</tr>
<tr>
<td></td>
<td>All (women)</td>
<td>GECCO</td>
<td>6132</td>
<td>6380</td>
<td>0.75/1.33</td>
<td>0.80/1.25</td>
</tr>
<tr>
<td></td>
<td>All (men)</td>
<td>GECCO</td>
<td>5356</td>
<td>5297</td>
<td>0.73/1.37</td>
<td>0.78/1.28</td>
</tr>
<tr>
<td></td>
<td>Colon</td>
<td>GECCO</td>
<td>7678</td>
<td>11679</td>
<td>0.78/1.28</td>
<td>0.83/1.20</td>
</tr>
<tr>
<td></td>
<td>Rectal</td>
<td>GECCO</td>
<td>2783</td>
<td>11679</td>
<td>0.68/1.47</td>
<td>0.75/1.33</td>
</tr>
<tr>
<td></td>
<td>Distal colon</td>
<td>GECCO</td>
<td>3354</td>
<td>11679</td>
<td>0.70/1.43</td>
<td>0.77/1.30</td>
</tr>
<tr>
<td></td>
<td>Proximal colon</td>
<td>GECCO</td>
<td>4185</td>
<td>11679</td>
<td>0.73/1.37</td>
<td>0.79/1.27</td>
</tr>
<tr>
<td>Breast</td>
<td>All</td>
<td>DRIVE</td>
<td>15748</td>
<td>18084</td>
<td>0.84/1.19</td>
<td>0.87/1.15</td>
</tr>
<tr>
<td></td>
<td>ER−</td>
<td>DRIVE</td>
<td>4939</td>
<td>13128</td>
<td>0.75/1.29</td>
<td>0.80/1.22</td>
</tr>
<tr>
<td>Prostate</td>
<td>All</td>
<td>PRACTICAL</td>
<td>22898</td>
<td>23054</td>
<td>0.86/1.16</td>
<td>0.89/1.12</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>GAME-ON</td>
<td>14159</td>
<td>12712</td>
<td>0.82/1.22</td>
<td>0.86/1.17</td>
</tr>
<tr>
<td></td>
<td>Aggressive</td>
<td>GAME-ON</td>
<td>4445</td>
<td>12724</td>
<td>0.74/1.30</td>
<td>0.79/1.23</td>
</tr>
<tr>
<td>Ovarian</td>
<td>All</td>
<td>FOCI</td>
<td>4369</td>
<td>9123</td>
<td>0.73/1.33</td>
<td>0.79/1.25</td>
</tr>
<tr>
<td></td>
<td>Clear-cell</td>
<td>FOCI</td>
<td>356</td>
<td>9123</td>
<td>0.19/1.86</td>
<td>0.36/1.67</td>
</tr>
<tr>
<td></td>
<td>Endometrioid</td>
<td>FOCI</td>
<td>715</td>
<td>9123</td>
<td>0.43/1.62</td>
<td>0.55/1.48</td>
</tr>
<tr>
<td></td>
<td>Serous</td>
<td>FOCI</td>
<td>2556</td>
<td>9123</td>
<td>0.67/1.39</td>
<td>0.74/1.30</td>
</tr>
<tr>
<td>Lung</td>
<td>All</td>
<td>TRICL-ILCCO</td>
<td>12537</td>
<td>17285</td>
<td>0.83/1.20</td>
<td>0.86/1.16</td>
</tr>
<tr>
<td></td>
<td>Adenocarcinoma</td>
<td>TRICL-ILCCO</td>
<td>3804</td>
<td>16289</td>
<td>0.73/1.30</td>
<td>0.78/1.23</td>
</tr>
<tr>
<td></td>
<td>Squamous</td>
<td>TRICL-ILCCO</td>
<td>3546</td>
<td>16434</td>
<td>0.72/1.31</td>
<td>0.78/1.24</td>
</tr>
<tr>
<td>Pancreatic</td>
<td>All</td>
<td>PanScan1‡</td>
<td>1896</td>
<td>1939</td>
<td>0.59/1.67</td>
<td>0.67/1.69</td>
</tr>
<tr>
<td>Neuroblastoma</td>
<td>All</td>
<td>Capasso, et al17‡</td>
<td>1627</td>
<td>3254</td>
<td>0.58/1.57</td>
<td>0.66/1.43</td>
</tr>
</tbody>
</table>

NR=not reported; ER=estrogen receptor negative.

*Minimum detectable odds ratio per 1 SD increase/decrease in 25(OH)D concentration; assume 80% power, 5% alpha level, and that 3% or 5% of 25(OH)D variance is explained by four single nucleotide polymorphisms (rs2282679, rs10741657, rs12785878, rs6013897) used in this paper. 1 SD in 25(OH)D corresponds to about 25 nmol/L (10 ng/mL).

1Summary random effects odds ratio and 95% confidence intervals for association of continuous 25(OH)D concentration (per 25 nmol/L) and risk of cancer in most recent published meta-analysis that reported dose-response summary result.

‡Obtained through MR-base platform.
25(OH)D concentration to risk of cancer. Similar minimum detectable effect sizes were estimated for cancer subtypes, except for clear cell and endometrioid ovarian carcinomas, for which there was adequate power to detect only large effects (for example, odds ratios 0.19-0.43). The power was larger if we assumed that 5% of the 25(OH)D variance was explained by the single nucleotide polymorphisms (table 1). Table 2 shows information on the associations of rs2282679, rs10741657, rs12785878 and rs6013897 with 25(OH)D concentration.

Association between individual single nucleotide polymorphisms and cancer

Appendix 2 shows the association between each single nucleotide polymorphism related to vitamin D and risk of colorectal, breast, prostate, ovarian, lung, and pancreatic cancer and neuroblastoma and their subtypes with data from GAME-ON, GECCO, PRACTICAL, and the MR-Base platform. None of the four polymorphisms was significantly associated with any cancer risk, except for rs6013897 with prostate cancer in the GAME-ON data (odds ratio per effect allele 1.06, 95% confidence interval 1.01 to 1.10; P=0.02), but this association was not observed in the larger PRACTICAL data (1.00, 0.97 to 1.04; P=0.81). The rs6013897 polymorphism was also associated with risk of colon cancer in the GECCO data (0.94, 0.89 to 0.99; P=0.03).

Mendelian randomisation estimates for multi-single nucleotide polymorphism scores

Based on Mendelian randomisation analyses with either the inverse variance weighted method or the

---

**Table 2 | Characteristics of genetic variants associated with 25(OH)D concentration in published genome-wide association studies**

<table>
<thead>
<tr>
<th>Single nucleotide polymorphism</th>
<th>Chromosome</th>
<th>Locus</th>
<th>Risk allele</th>
<th>β estimate†</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2282679</td>
<td>4</td>
<td>GC</td>
<td>G</td>
<td>−4.67</td>
<td>&lt;3.4×10−302</td>
</tr>
<tr>
<td>rs10741657</td>
<td>11</td>
<td>CYP2R1</td>
<td>G</td>
<td>−1.72</td>
<td>6.5×10−93</td>
</tr>
<tr>
<td>rs12785878</td>
<td>11</td>
<td>DHCR7/NADSYN1</td>
<td>G</td>
<td>−2.11</td>
<td>6.4×10−135</td>
</tr>
<tr>
<td>rs6013897</td>
<td>20</td>
<td>CYP24A1</td>
<td>A</td>
<td>−0.98</td>
<td>3.4×10−12</td>
</tr>
</tbody>
</table>

*Source: Vimaleswaran, et al, 2013.†Reported per unit change in nmol/L in natural scale per effect allele.

---

**Table 3 | Mendelian randomisation estimates between multi-single nucleotide polymorphism risk scores of continuous 25(OH)D and risk of cancer calculated with inverse variance weighted method and likelihood method**

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Study</th>
<th>OR* (95% CI); P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>GAME-ON</td>
<td>1.04 (0.78 to 1.38); 0.81</td>
</tr>
<tr>
<td>All</td>
<td>GECCO</td>
<td>0.92 (0.76 to 1.10); 0.36</td>
</tr>
<tr>
<td>All (women)</td>
<td>GECCO</td>
<td>0.92 (0.71 to 1.18); 0.52</td>
</tr>
<tr>
<td>All (men)</td>
<td>GECCO</td>
<td>0.91 (0.70 to 1.20); 0.52</td>
</tr>
<tr>
<td>Colon</td>
<td>GECCO</td>
<td>0.90 (0.73 to 1.11); 0.33</td>
</tr>
<tr>
<td>Rectal</td>
<td>GECCO</td>
<td>0.93 (0.68 to 1.26); 0.64</td>
</tr>
<tr>
<td>Distal colon</td>
<td>GECCO</td>
<td>0.97 (0.73 to 1.28); 0.83</td>
</tr>
<tr>
<td>Proximal colon</td>
<td>GECCO</td>
<td>0.83 (0.64 to 1.07); 0.14</td>
</tr>
<tr>
<td>Breast</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>DRIVE</td>
<td>1.05 (0.89 to 1.24); 0.59</td>
</tr>
<tr>
<td>ER−</td>
<td>DRIVE</td>
<td>1.15 (0.88 to 1.50); 0.30</td>
</tr>
<tr>
<td>Prostate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>PRACTICAL</td>
<td>0.89 (0.77 to 1.02); 0.08</td>
</tr>
<tr>
<td>All</td>
<td>GAME-ON</td>
<td>1.08 (0.88 to 1.33); 0.47</td>
</tr>
<tr>
<td>Aggressive</td>
<td>GAME-ON</td>
<td>1.14 (0.85 to 1.54); 0.38</td>
</tr>
<tr>
<td>Ovarian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>FOCI</td>
<td>1.12 (0.86 to 1.47); 0.40</td>
</tr>
<tr>
<td>Clear-cell</td>
<td>FOCI</td>
<td>0.99 (0.66 to 1.49); 0.98</td>
</tr>
<tr>
<td>Endometrioid</td>
<td>FOCI</td>
<td>0.83 (0.48 to 1.43); 0.51</td>
</tr>
<tr>
<td>Serous</td>
<td>FOCI</td>
<td>1.26 (0.91 to 1.76); 0.17</td>
</tr>
<tr>
<td>Lung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>TRICL-ILCCO</td>
<td>1.03 (0.87 to 1.23); 0.72</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>TRICL-ILCCO</td>
<td>1.03 (0.79 to 1.35); 0.84</td>
</tr>
<tr>
<td>Squamous</td>
<td>TRICL-ILCCO</td>
<td>0.95 (0.72 to 1.25); 0.74</td>
</tr>
<tr>
<td>Pancreatic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>PanScan1†</td>
<td>1.36 (0.81 to 2.27); 0.25</td>
</tr>
<tr>
<td>Neuroblastoma</td>
<td></td>
<td>1.36 (0.80 to 2.27); 0.25</td>
</tr>
<tr>
<td>All</td>
<td>Capasso, et al17†</td>
<td>0.76 (0.47 to 1.21); 0.24</td>
</tr>
</tbody>
</table>

ER−=oestrogen receptor negative.

*Represents increase/decrease of risk per 25 nmol/L increase in nmol/L in natural scale of 25(OH)D. All four single nucleotide polymorphisms were used for all cancers, except for pancreatic cancer and neuroblastoma, for which only two polymorphisms (rs10741657, rs2282679) were available.

†Obtained through MR-base platform.
likelihood based method, we found little evidence that the multi-polymorphism scores for continuous 25(OH)D concentration were associated with risk of colorectal, breast, prostate, ovarian, lung, or pancreatic cancer and neuroblastoma or their subtypes (table 3). Figures 1-6 show scatter plots of associations between vitamin D polymorphism and risk of various types of cancer. Plots are overlaid by Mendelian randomisation estimate (slope of solid line) and its 95% confidence interval (dotted lines) of multi-polymorphism score of continuous circulating 25(OH)D on risk of the seven cancers and their subtypes. We found a marginally significant association for total prostate cancer, for which a genetically determined 25 nmol/L increase in 25(OH)D concentration yielded an odds ratio of 0.89 (95% confidence interval 0.77 to 1.02; P=0.08; fig 3).

Assessment of Mendelian randomisation assumptions
Mendelian randomisation estimates have a causal interpretation only if the instrumental variable assumptions of the method are valid. To satisfy the first assumption, we selected single nucleotide polymorphisms with a genome-wide significant

Fig 1 | Association between single nucleotide polymorphisms associated with vitamin D and risk of colorectal cancer and circulating 25(OH)D concentration. Per allele associations with risk plotted against per allele associations with continuous circulating 25(OH)D concentration (vertical and horizontal black lines around points show 95% confidence interval for each polymorphism)
We obtained estimates of association with continuous 25(OH)D concentration for each polymorphism from a previous large Mendelian randomisation study, which estimated that the F statistic was 230 (n=35873) and 489 (n=38191) for the vitamin D synthesis and metabolism allele scores, respectively.28

We carried out statistical tests and sensitivity analyses to evaluate the potential violation of the second and third assumptions. The goodness of fit tests indicated absence of horizontal pleiotropic effects of the four polymorphisms on cancer that are unrelated to the effect of each polymorphism on circulating 25(OH)D (table A in appendix 3). Over-identification tests also suggested that the effect estimates with different genetic variants were similar for all cancers. When we further evaluated presence of horizontal pleiotropy by performing the Mendelian randomisation analysis using two separate allelic scores (that is, vitamin D synthesis and metabolism), the results were identical and non-significant for all cancers (tables B and C in appendix 3). The MR-Egger regression method also did not show any evidence for the presence of horizontal pleiotropy for any of the reported associations (table D in appendix 3), as the P values for the intercept were large and the estimates adjusted for pleiotropy suggested null effects, although this method is expected to have low power to detect violation of assumptions when only four genetic instruments are used. The weighted median method also yielded no significant estimates (table D in appendix 3). We found no evidence in published genome-wide association studies that the four single nucleotide polymorphisms associated with vitamin D were genome-wide significantly associated with any other phenotype except 25(OH)D concentrations, which means that the third Mendelian randomisation assumption is probably not violated. Additionally, previous Mendelian randomisation studies using individual level data found no evidence for association between the vitamin D polymorphisms and potential environmental confounders.9 42 43
Main findings and comparisons with the literature

In this large Mendelian randomisation study, we observed little evidence that a multi-single nucleotide polymorphism score for circulating 25(OH)D concentration was associated with risk of several cancers, including colorectal, breast, prostate, ovarian, lung, and pancreatic cancer and neuroblastoma or some of their subtypes. This was the first study with sufficient sample size under Mendelian randomisation assumptions to show a lack of causal effect for a linear association between 25(OH)D concentration and risk of these cancers.

**Colorectal cancer**

The overall evidence for an association between vitamin D and risk of specific cancers is mixed. Higher circulating 25(OH)D concentration has been associated with a lower risk of colorectal cancer. A systematic review of eight prospective studies that included 2690 cases of colorectal cancer observed a 34% (odds ratio 0.66, 95% confidence interval 0.54 to 0.81) lower risk of colorectal cancer for the top compared with the bottom quartile of 25(OH) D concentration. Another meta-analysis found that a 25 nmol/L increment in circulating 25(OH)D was associated with a relative risk of 0.74 (0.63 to 0.89). Although the current Mendelian randomisation study was powered for minimum detectable odds ratios up to 0.85 per 25 nmol/L in circulating 25(OH)D, it did not support such an association. In agreement with our findings, a previous Mendelian randomisation analysis of a Scottish case-control study of 2001 cases and 2237 controls did not find an association between genetically determined vitamin D concentrations and risk of colorectal cancer. The GECCO consortium found no association between the four single nucleotide polymorphisms associated with vitamin D and risk of colorectal cancer.

**Breast cancer**

The prospective epidemiological evidence for an association between circulating 25(OH)D concentrations and risk of breast cancer is inconclusive. A meta-analysis by Gandini and colleagues reported that a 25 nmol/L increment in circulating 25(OH)D concentration was associated with a relative risk of 0.89 (95% confidence interval 0.81 to 0.98). Two subsequent meta-analyses observed no association between 25(OH)D concentration and risk of premenopausal breast cancer, whereas an inverse association was suggested for postmenopausal breast cancer. Specifically, Bauer and colleagues suggested a borderline significant inverse association for postmenopausal breast cancer with a relative risk per 12.5 nmol/L of circulating 25(OH)D of 0.97 (0.93 to 1.00). We did not find an association between genetically determined 25(OH)D concentrations and risk of breast cancer, and our study was powered to find minimum detectable odds ratios ranging from 0.84 to 0.87 per 25 nmol/L in 25(OH)D. Information on menopausal status was not available in the large genetic networks that we used, but most women in our sample had postmenopausal breast cancer. In agreement with our findings, the Women's Health Initiative trial of vitamin D plus calcium supplementation in postmenopausal...
women did not support a protective association with breast cancer (hazard ratio 0.96, 95% confidence interval 0.85 to 1.09). A large cohort consortium of 9456 cases and 10816 controls also found no association between the four single nucleotide polymorphisms associated with vitamin D and risk of breast cancer.

**Prostate cancer**

A meta-analysis of 14 prospective studies published in 2011 provided little evidence that 25(OH)D concentration was associated with risk of total (odds ratio per 25 nmol/L, 1.04, 95% confidence interval 0.99 to 1.10) or aggressive (0.98, 0.84 to 1.15) prostate cancer. More recent prospective studies have reported null associations between circulating 25(OH)D concentration and risk of total prostate cancer, but inverse associations for aggressive or lethal disease.

Other prospective studies have reported positive associations for total disease and null associations for lethal disease, or a significant U shaped association for total and aggressive disease; whereas, a meta-analysis of 17 prospective studies published in 2014 observed a significantly increased risk of total prostate cancer (relative risk 1.18, 95% confidence interval 1.07 to 1.30) for the highest compared with the lowest concentrations of circulating 25(OH)D. A large cohort consortium of 10 018 cases of total prostate cancer and 11 052 controls found a significant association between a genetic risk score of the four single nucleotide polymorphisms associated with vitamin D and the risk of aggressive, but not total, prostate cancer. Our Mendelian randomisation analysis of 22 898 cases and 23 054 controls found no strong evidence for an association between genetically determined circulating 25(OH)D concentrations and risk of total or aggressive prostate cancer.

**Ovarian cancer**

Few prospective epidemiological studies have examined the association between circulating 25(OH)D concentrations and risk of ovarian cancer, and most have yielded null results. A recent Mendelian randomisation study by Ong and colleagues observed a significant decrease (odds ratio 0.988, 95% confidence interval 0.979 to 0.997) in risk of ovarian cancer per 1 nmol/L increase in 25(OH)D concentrations, which
translates to an odds ratio of 1.27 (1.06 to 1.51) per 20 nmol/L decrease in 25(OH)D. This study was twice the size of our Mendelian randomisation study for the gene-outcome associations but used three instead of four single nucleotide polymorphisms.\textsuperscript{57} Additionally, the authors used published estimates for the association between rs2282679 and 25(OH)D concentrations from a small cohort of 2347 participants,\textsuperscript{43} whereas our analysis used published estimates from a large meta-analysis of about 38 000 participants.\textsuperscript{28} We did not find a significant association (odds ratio per unit increase in 25(OH)D, 1.005, 0.994 to 1.016). When we re-ran the Mendelian randomisation study using three (rs2282679, rs10741657, rs12785878) instead of four polymorphisms or using the same estimate for the association between rs2282679 with 25(OH)D concentrations as in the paper by Ong and colleagues, we observed almost identical non-significant results. Therefore, the small difference between the two Mendelian randomisation studies is plausible and can be explained by the larger statistical power of the gene-outcome association in the previous study.

Lung and pancreatic cancer and neuroblastoma

There is limited epidemiological evidence for a role of vitamin D in risk of lung and pancreatic cancer and neuroblastoma. A meta-analysis of 10 prospective studies reported a significant reduction (relative risk 0.95, 95% confidence interval 0.91 to 0.99) in risk of lung cancer for each 10 nmol/L increment in 25(OH)D concentration, but the heterogeneity between studies was large and a potential non-linear relation was suggested.\textsuperscript{41} We did not observe a significant association between genetically determined 25(OH)D concentrations and risk of lung or pancreatic cancer or neuroblastoma in the current Mendelian randomisation study, but our study was not powered to detect the small effect sizes suggested by the published meta-analysis for lung cancer.

Strengths and limitations of this study

The main benefit of Mendelian randomisation studies is that they avoid biases that are commonly present in conventional observational literature. Resulting estimates have a causal interpretation only if the assumptions of the method hold. Though it is not possible to prove the validity of the assumptions, we performed sensitivity analyses and used several statistical tests to look for potential violations. We found no evidence of violation, though some of the statistical tests have low power to detect this when few genetic instruments are used (for example, MR-Egger).\textsuperscript{34} Previous Mendelian randomisation studies on vitamin D and risk of cancer or death that used individual level data, however, also did not suggest any violation of assumptions.\textsuperscript{9,42} We used summary data for seven cancers and several of their subtypes, using thousands of cases of cancer and controls from several large genetic consortiums and published genome-wide association studies. We were powered to detect effect sizes of moderate magnitude for most primary cancer outcomes, but we cannot exclude the existence of causal clinically relevant effects of low magnitude.

Several limitations should be also considered in interpreting our findings. The summary level data that we used did not allow for stratified analyses by covariates of interest, such as age, sex, menopausal status, smoking, body mass index (BMI), and use of hormone replacement therapy or by other related genes or according to whether populations were vitamin D deficient or not. In addition, we could not explore potential non-linear associations between 25(OH)D concentrations and risk of cancer, which have been suggested by some studies. Furthermore, the currently known single nucleotide polymorphisms associated with vitamin D account for only a small amount of the variance observed in 25(OH)D concentration, but previous Mendelian randomisation studies have identified significant associations between vitamin D and several outcomes.\textsuperscript{9,43,57,58} In addition, these single nucleotide polymorphisms do not predict concentrations of 1,25-dihydroxyvitamin D, which is the most biologically active metabolite of vitamin D, and also cannot predict vitamin D concentrations at the cellular level. Therefore, our results cannot be considered definitive. Future large pooling consortia, larger genome-wide association studies of 25(OH)D concentration, and Mendelian randomisation studies with individual level data could deal with the latter issues. Moreover, large scale, general population, high dose vitamin D supplementation trials designed to overcome many of the limitations of previous trials (such as modest size, inadequate dose, relatively short duration, and small number of cancers) are ongoing\textsuperscript{59,60} and might provide an improved understanding on the role of supplementation for development and death from non-skeletal outcomes.

Conclusion

In summary, using a comprehensive Mendelian randomisation study, we found little evidence for linear causal associations between genetic determinants of circulating vitamin D concentration and risk of colorectal, breast, prostate, ovarian, lung, and pancreatic cancer and neuroblastoma, but we cannot rule out the existence of causal clinically relevant effects of low magnitude. Our results, in combination with previous literature, provide evidence that population-wide screening for vitamin D deficiency and subsequent widespread vitamin D supplementation should not currently be recommended as a strategy for primary cancer prevention.

AUTHOR AFFILIATIONS

1 Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina, Greece
2 School of Mathematics and Statistics, University College Dublin, Dublin, Ireland
3 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK
4 School of Social and Community Medicine, University of Bristol, Bristol, UK
5 MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
Members of the collaboration: Vasiliki Dimitrakopoulou (Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina, Greece; School of Mathematics and Statistics, University College Dublin, Dublin, Ireland); Konstantinos K Tsilidis (Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina, Greece; Department of Food and Nutritional Sciences, Hugh Sinclair Unit of Human Nutrition and Institute for Cardiovascular and Metabolic Research (ICMR), University of Reading, Reading, UK); Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; Department of Public Health Sciences, University of Chicago, Chicago, IL, USA; Comprehensive Cancer Center, University of Chicago, Chicago, IL, USA; Department of Human Genetics, University of Chicago, Chicago, IL, USA; Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA.

Members of the collaboration: Stephanie Stenning (Department of Radiology, University Hospital of Zurich, Zurich, Switzerland); Maxime Bihist (Department of Radiology, University of Montreal, Montreal, Canada); Richard LeMarchand (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, US); Christopher I Amos (Department of Preventive Medicine, University of Southern California, Los Angeles, CA, US); Victoria Stevens (Department of Preventive Medicine, Washington State University, Pullman, WA, US); Christopher A Haiman (Department of Preventive Medicine, University of Southern California, Los Angeles, CA, US); John A Baron (Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC, US); Peter T Campbell (Epidemiology Program, American Cancer Society, Atlanta, GA, US); Neil Caporaso (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, US); Christopher I Amos, Department of Biomedical Data Sciences, Geisel School of Medicine at Dartmouth, Lebanon, NH, US; Xiafeng Wu (University of Texas MD Anderson Cancer Center, Houston, TX, US); Rayjean J Hung (Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada); Heike Bickeboller (Department of Genetic Epidemiology, University Medical Centre, Georg-August-Universität Göttingen, Göttingen, Germany); Angela Rise (Division of Cancer Epidemiology and Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany); Stefan Edeiken (Institute of Medical Informatics, Biometry and Epidemiology, Ludwig Maximilians University, Munich, Germany); Michael Hoffmeister (Division of Clinical Epidemiology and Biostatistics, German Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany); Richard Houlston (Division of Genetics and Epidemiology, Institute of Cancer Sciences, Imperial College London, London, UK); Richard Houlston (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, US); Christopher I Amos (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, US); Victoria Stevens (Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, US).


Appendix 1: Supplementary methods
Appendix 2: Forest plots of associations
Appendix 3: Supplementary tables
Appendix 4: Further details of contributing consortiums, funding, and acknowledgments