Imperial College London

Vitamin D exposure and cancer incidence and mortality, all-cause mortality, and cause-specific mortality in the European Prospective Investigation into Cancer and Nutrition (EPIC)

Neil Murphy
Submitted for the degree of Doctor of Philosophy
Department of Epidemiology & Biostatistics
School of Public Health
School of Medicine
Imperial College London

DECLARATION OF ORIGINALITY

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any all material which is not my own work has been properly acknowledged.

COPYRIGHT DECLARATION

ACKNOWLEDGEMENTS

ABSTRACT

Despite extensive epidemiological research and plausible biological mechanisms being elucidated, it is unclear whether vitamin D reduces risks of cancer incidence and mortality. Only for colorectal cancer does the observational evidence seem persuasive, whereas for other cancer types an anti-carcinogenic role has not been established convincingly, with rarer cancers seldom investigated. Similarly, whether vitamin D has a beneficial role on other chronic disease end-points and all-cause mortality remains uncertain, despite extensive research.

Prospective studies which directly measure actual circulating 25-hydroxyvitamin D (25(OH)D) are viewed as the "gold standard" approach to assess vitamin D-disease associations. However, these studies are expensive to carry out (as circulating 25(OH)D usually has to be measured in all participants) and a single measurement of circulating 25(OH)D may not reflect long-term exposures (due to within-person variability). An alternative approach, not yet used in European populations, is to create predictor scores of circulating 25(OH)D levels. This cost effective approach provides the opportunity to examine associations between predicted 25(OH)D and multiple outcomes (including less common diseases).

Sex-specific predictor scores were derived in 4,089 participants from the European Investigation into Cancer and Nutrition (EPIC) study by quantifying the relationships between correlates/determinants of circulating 25(OH)D levels (using multivariable linear regression models). The predictor scores were validated in 2,029 participants with measured circulating 25(OH)D levels. In summary, the predictor scores provided poor estimates of absolute circulating 25(OH)D levels but were more successful at ranking individuals similarly by their actual and predicted levels. The predictor scores were also able to replicate results from previous EPIC colorectal cancer incidence and prostate cancer incidence nested case-control studies which used actual circulating 25(OH)D measurements. Overall, this evidence suggests that the predictor scores may have utility for epidemiological research but not in a clinical setting.

The predictor scores were then applied to the full EPIC cohort to assess the associations between predicted 25(OH)D levels with risks of cancer incidence and mortality, all-cause mortality, and cause-specific mortality. In summary, significant inverse predicted 25(OH)D score associations were observed for: overall cancer incidence and mortality; colorectal cancer incidence; lung cancer incidence and mortality; kidney cancer incidence; stomach and oesophageal cancer incidence; pancreatic cancer incidence and mortality; thyroid cancer incidence; prostate cancer mortality; all-cause mortality; circulatory disease mortality; respiratory disease mortality; and digestive disease mortality. However, due to the methodological limitations specific to 25(OH)D predictor scores - such as providing poor estimates of absolute levels - and observational epidemiology in general, it is important to acknowledge that alternative explanations may explain some or all of these observed relationships.

CONTENTS

TABLES AND FIGURES

ABBREVIATIONS

$1,25(OH)_2D$	1α, 25-dihydroxyvitamin D_3
25(OH)D	25-hydroxyvitamin D
95% CI	95% Confidence interval
ATBC	The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study
BMI	Body mass index (kg/m^2)
CPS-II	Cancer Prevention Study-II Nutrition Cohort
EPIC	The European Prospective Investigation into Cancer and Nutrition
HPFS	Health Professionals Follow-up Study
HR	Hazard ratio
NCC	Nested case-control study
NHANES III	National Health and Nutrition Examination Survey in the USA III
NHS	Nurses' Health Study in the USA
OR	Odds ratio
PLCO	Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial
PTH	Parathyroid hormone
RCT	Randomised controlled trial
RR	Relative risk
SD	Standard deviation
SPF	Sun protection factor
UVB	Ultraviolet B
VDP	Vitamin D binding protein
VDPP	The Vitamin D Pooling Project of Rarer Cancers
VDR	Vitamin D receptor
WHI	Women's Health Initiative

1. INTRODUCTION

Vitamin D is predominantly known for its role in calcium and phosphorus metabolism but a potential anti-carcinogenic role was first proposed by Garland and Garland in 1980 (1). This hypothesis was further developed by the identification of the vitamin D receptor (VDR) and 1α-hydroxylase (the enzyme which converts vitamin D into its most active form) throughout the body. Despite extensive epidemiological research and plausible biological mechanisms being elucidated, it is unclear whether vitamin D reduces risks of cancer incidence and mortality. Only for colorectal cancer does the observational evidence seem persuasive, whereas for other cancer types an anti-carcinogenic role has not yet been established convincingly. Possible biological roles for vitamin D in other chronic disease end-points have also been investigated. For circulatory disease mortality, inverse relationships have usually been reported; however, to date, only relatively small studies have been conducted. For respiratory disease and digestive disease mortality, little is known about a possible role for vitamin D as minimal prospective studies have been carried out. For all-cause mortality, previous research is suggestive of elevated risk at lower levels of circulating vitamin D; although recent research suggests a J- or U-shaped risk curve may exist, with increased risk observed at higher levels of circulating vitamin D. Thus, despite the high volume of research, many questions on the role of vitamin D on cancer incidence and mortality, circulatory disease, respiratory disease, digestive disease and all-cause mortality remain unanswered.

In the first part of this analysis the relationships between dietary vitamin D intake and all-cause and cancer caused mortality within the European Prospective Investigation into Cancer and Nutrition (EPIC) were investigated. In the second part of this study, data from EPIC were used to derive and validate predictor scores of circulating vitamin D status. These predictor scores was then applied to the full EPIC cohort to assess risks of cancer incidence and mortality, and all-cause mortality, circulatory disease mortality, respiratory disease mortality, and digestive disease mortality.

1.1 Vitamin D

1.1.1 Production and metabolism of vitamin D

Vitamin D is a fat soluble secosteroid which can be produced endogenously or obtained exogenously from dietary or supplementary sources (Figure 1). The two major forms are vitamin D_2 (ergocalciferol; $C_{28}H_{44}O$) and D_3 (cholecalciferol; $C_{27}H_{44}O$). Endogenously produced Vitamin D_3 is produced by UVB (ultraviolet B) radiation (wavelength 290-315 nm) converting 7-dehydrocholesterol into previtamin D_3. An immediate thermally induced isomerisation changes previtamin D_3 into vitamin D_3. Vitamin D-binding protein then transports vitamin D_3 into the liver. Exogenous dietary vitamin D exists as either vitamin D_2 or D_3. After ingestion these fat soluble compounds are incorporated into chylomicrons, absorbed through the lymphatic system, and then transported to the liver. In the liver, both endogenous and exogenous vitamin D are converted into 25-hydroxyvitamin D (25(OH)D; $C_{27}H_{44}O_2$) in a reaction catalysed by 25-hydroxylase. The biologically active form of the vitamin is 1α,25-dihydroxyvitamin D (1,25(OH)$_2$D; $C_{27}H_{44}O_3$), and its formation from 25(OH)D is catalysed by the enzyme 1α-hydroxylase. This occurs within the kidney (where it is dependent on calcium and parathyroid hormone (PTH) concentrations) and in cells throughout the body. Circulating 25(OH)D is the main circulating form of vitamin D and is used as an indicator of body status as its half-life is estimated to be three weeks (2). In contrast, 1,25(OH)$_2$D has an estimated half-life of less than four hours (3).

The biological actions of 1,25(OH)$_2$D are achieved by binding to the VDR, an intracellular hormone receptor, and can be categorised into genomic and rapid responses (4). The genomic response is the better understood. Once activated by 1,25(OH)$_2$D, the VDR interacts with the retinoid X receptor, which in turn binds with 1,25(OH)$_2$D (5;6). The activated VDR then modulates gene transcription by binding up or down to vitamin D responsive elements (5;6). Over 2,700 human genome sites have been identified as being involved in VDR binding (7). A recent genome wide study identified genes which VDR bind to that are associated with various health conditions, including: Crohn's disease, type-1 diabetes, and colorectal cancer (7). The expression of up to 229 genes may be affected by 1,25(OH)$_2$D (6;7). Rapid or non-genomic responses of vitamin D have recently been identified. Once more, 1,25(OH)$_2$D actions are mediated through VDR; except this time the VDR are membrane bound and associated with caveolae domains, rather than located within the nucleus (4). Once membrane bound VDR have been activated, the rapid

responses are instigated through secondary messengers. This response has been proposed to explain the function of vitamin D in influencing intestinal absorption of calcium, insulin secretion of pancreatic β cells, and calcium entering muscle cells (6).

The VDR is a vital mediator for the cellular effects of vitamin D. The gene encoding VDR is located at chromosome 12q13.11. To date, more than 60 polymorphisms of the VDR gene have been identified (8), but only a few VDR single nucleotide polymorphisms have been extensively studied with regard to cancer and other chronic diseases (9). One of these is the rs11568820 polymorphism (aka Cdx-2), which is situated in exon 1e and modulates transcription of VDR gene expression. Another is polymorphism rs10735810 (aka FokI) which is located in the coding region of the VDR gene and therefore has an effect on the activity of the receptor. Finally the rs1544410 (aka Bsml), rs7675232 (aka Apal) and rs731236 (aka Taq1) single nucleotide polymorphisms are located in the 3' end of the VDR gene. These polymorphisms have been hypothesised to alter the risks of colorectal cancer, breast cancer and prostate cancer. However, a recent review reported that these VDR polymorphisms did not substantially influence disease risk (10). Larger studies which can investigate how environmental exposures interact with these polymorphisms are required to broaden the understanding of how genetic variation may influence disease risk.

Figure 1. Vitamin D production and metabolism
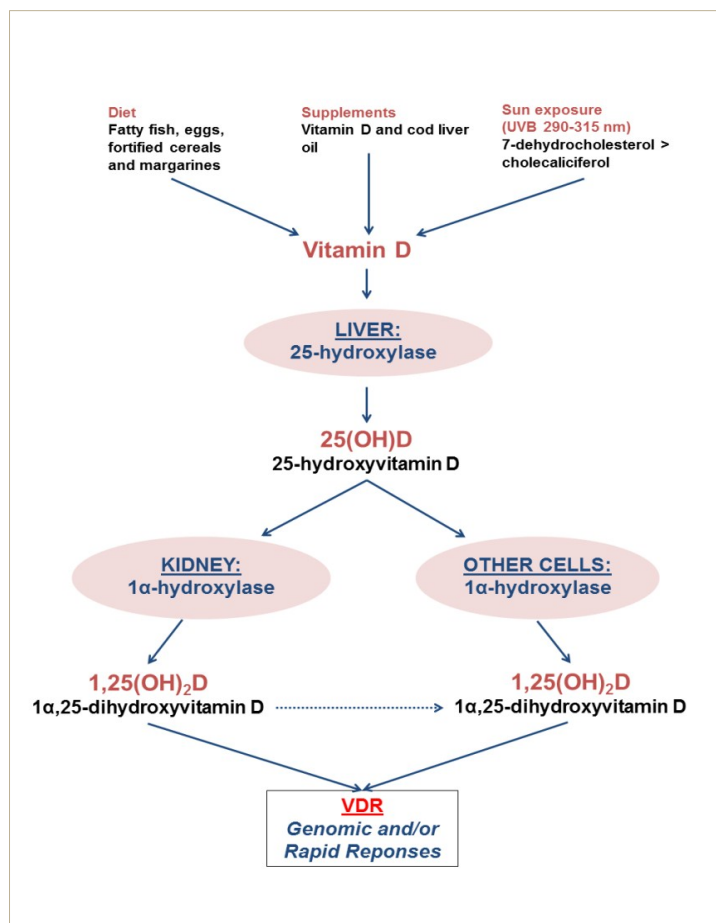
1.1.2 Biological roles of vitamin D

Skeletal roles

The skeletal roles of vitamin D are well defined, relative to the non-skeletal roles. In summary, vitamin D acts upon the intestine and bone to ensure calcium and phosphorus homeostasis (Figure 2). In the intestine, $1,25(OH)_2D$ promotes the absorption of calcium and phosphorus. Specifically for calcium, this involves the up regulation of the expression of the epithelial calcium channel and calcium binding protein (6). In the bone, $1,25(OH)_2D$ stimulates resorption by increasing the number of osteoclasts; this results in an increase of calcium and phosphorus into the circulation.

Figure 2. The summarised biological roles of Vitamin D related to chronic diseases
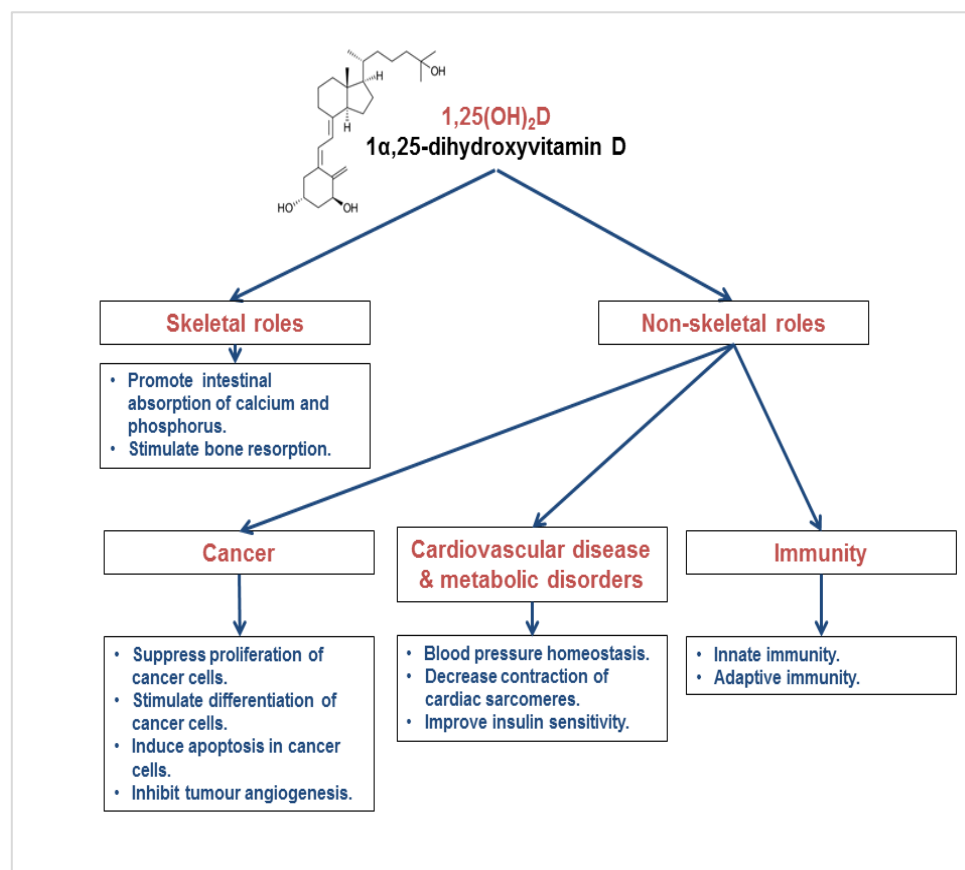
Non-skeletal roles

The biological plausibility of non-skeletal roles for vitamin D has grown from the identification of the VDR and 1-α-hydroxylase throughout the body. VDR expression has been identified in at least 38 different tissue types, including: adipose, brain, breast cancer cells, colon, liver, lung, muscle, ovary, pancreas, prostate, skin, stomach, and thyroid (4;7). Expression of 1α-hydroxylase, not influenced by PTH and calcium concentration, has been identified in cancer and non-cancer cells, such as prostate, colon, lung, endothelial, brain, pancreatic β-cells, and monocytes (4;11). Outlined below are proposed effects of vitamin D on cancer, immunity, cardiovascular diseases and metabolic disorders (Figure 2).

Cancer

Laboratory research suggests that vitamin D and its metabolites may reduce incidence of many types of cancers by mechanisms said to include: suppression of proliferation and stimulating differentiation in cancer cells; inducing apoptosis in cancer cells; and inhibiting tumour angiogenesis.

Suppression of proliferation, and stimulation of differentiation in cancer cells

Separate anti-proliferative roles of vitamin D have been proposed. Firstly, $1,25(OH)_2D$ has been shown to promote cell cycle arrest by enhancing the expression of Cyclin-dependent kinase (CDK) inhibitors (6;12). For instance, elevated p21 and p27 gene expression was enhanced by $1,25(OH)_2D$ in squamous cancer cell lines within the head and neck (12). A proposed inhibitory mechanism of vitamin D against cellular proliferation is through aiding the preservation of gap junction intercellular communication during carcinogenesis; consequently strengthening contact inhibition of proliferation (13). Vitamin D has also been shown to promote differentiation within pathway target genes (14).

Inducing apoptosis in cancer cells

Experimental evidence has highlighted a regulatory role for vitamin D on the principal mediators of apoptosis in cancer cells (6;15). Specifically, $1,25(OH)_2D$ decreases the expression of BCL2 and BCL-XL (anti-apoptotic proteins) and increases the expression of BAX, BAK and BAD (pro-apoptotic proteins) (15). Apoptosis in tumour cells has also been shown to be constrained by the down regulation of telomerase activity by $1,25(OH)_2D$; leading to swifter telomere shortening (16).

Inhibition of tumour angiogenesis

In vitro and in vivo models have demonstrated that vitamin D has anti-angiogenic properties. In vitro, $1,25(OH)_2D$ has been shown to reduce vascular endothelial growth factor (VEGF) induced cell sprouting, elongation and proliferation (17). In vivo, tumour xenografts placed on mice, induced to overexpress VEGF, were treated with $1,25(OH)_2D$ and, after 8 weeks, reduced vascularisation was observed (17).

Immunity

Roles for vitamin D on the innate and adaptive immune systems have been discovered. For the innate immune system, $1,25(OH)_2D$ has been shown to induce the differentiation of

monocytes into macrophages (18). Furthermore, 1,25(OH)$_2$D has been shown to improve the chemotactic and phagocytic capabilities of macrophages (19). Antimicrobial actions of vitamin D, via 1,25(OH)$_2$D activation of the cathelicidin gene (CAMP) and expression of defensin β2, have also been uncovered (20). Vitamin D also influences the adaptive immunity response. Activation of the VDR can result in greater production of T regulatory cells by providing dendritic cells with their telerogenic properties (19). Also on T-cells, 1,25(OH)$_2$D has been shown to reduce the production of the Th1 (IL-2 and IFN-γ) and Th17 cytokines (6).

Cardiovascular disease and metabolic disorders

Vitamin D may be beneficial for cardiovascular health through blood pressure homeostasis, with two separate proposed mechanisms. Firstly, animal models have demonstrated that 1,25(OH)$_2$D is a negative regulator for the renin-angiotensin system (RAS), which regulates blood pressure, by reducing renin synthesis (21). Renin is a protease that hydrolyses angiotensinogen into angiotensin I, which in turn is converted into angiotensin II by the angiotension-converting enzyme. Angiotensin II elevates blood pressure by constricting blood vessels and increasing renal absorption of sodium and water (via aldosterone release) (6). A secondary mechanism whereby 1,25(OH)$_2$D lowers blood pressure is through the reduction of PTH levels, which have been associated with elevated blood pressure (6). A modulatory role for vitamin D on cardiac sarcomere contraction has also been proposed (22). Rat models have demonstrated a rapid response (within 2.5 minutes) for 1,25(OH)$_2$D on decreasing the contraction of sarcomeres (22).

A role for vitamin D on metabolic disorders, such as insulin resistance, has been proposed. Firstly, 1,25(OH)$_2$D treatment has been shown to promote transcription of U-937, the human promonocytic cell insulin resistance gene (23). Vitamin D may also indirectly influence insulin sensitivity through its role in calcium homeostasis. Calcium is required for insulin mediated responses in certain tissues, such as skeletal muscle and adipose tissue (24). Finally, vitamin D may also improve insulin sensitivity and promote β-cell survival through the modulation of cytokine levels; as elevated cytokine levels may trigger β-cell apoptosis (24).

1.1.3 Exogenous dietary sources of vitamin D (including supplements)

Natural dietary sources of vitamin D are limited. Good natural sources include fatty fish (such as mackerel and sardines) and egg yolks (Table 1). Fortification of margarine spreads and some cereal products with vitamin D_2 and D_3 mean they are also good sources. Vitamin D within foods remains stable through storage, processing, and cooking (25).

In the UK, dietary supplements usually contain 10-25 µg vitamin D_3 and are recommended by manufacturers to be taken daily. Cod liver oil is also a good source of vitamin D (Table 1). The contribution of dietary supplements to vitamin D status is country dependent. For instance, amongst Norwegian men and women, supplements contributed 42% and 49% of total dietary intake of vitamin D respectively (26). An analysis of 1958 British Birth Cohort, conducted in 2007, reported that 13% of men and 20% of women used vitamin D supplements (27).

Analyses on quantifying the relationship between intakes of vitamin D and circulating 25(OH)D has been conducted among men in winter months when endogenous synthesis is low (28). From a mean baseline of 70.3 nmol/L, a 1 µg increase in vitamin D_3 intake increased circulating 25(OH)D by 0.70 nmol/L. From a lower baseline level, 25(OH)D increases 1.2 nmol/L for a 1 µg increase in vitamin D_3 intake (29).

Table 1. Vitamin D content of foods (µg/100g)

(Sources: Olsen *et al.,* (25); SACN (30))

Milk and milk products	
Cow's milk	0.01-0.03
Human milk	0.04
Cream	0.1-0.3
Cheese	0.03-0.05
Yoghurt	Trace-0.04
Eggs	
Whole	1.8
Yolk	4.9
Fats and oils	
Butter	0.8
Cod liver oil	210
Fortified food items	
Breakfast cereals	3-8
Margarines and spreads	5.8-8.0
Meat and meat products	
Beef, lamb, pork, veal	Trace
Poultry, game	Trace
Liver	0.2-1.1
Fish and fish products	
White fish	Trace
Fatty fish	5-10
Crustacea and molluscs	Trace

1.1.4 Deficiency and toxicity

Deficiency

Deficiency of vitamin D impairs dietary calcium absorption and causes demineralisation of the skeleton. Severe deficiency is usually defined as circulating 25(OH)D levels below 25 nmol/L and at these levels rickets and osteomalacia can occur in children and adults. Circulating 25(OH)D levels between 25 and 50 nmol/L have been associated with elevated PTH, a biomarker of vitamin D insufficiency (9). Circulating 25(OH)D equal to or above 75 nmol/L has been quantified as the optimal level for bone health (29). At this level PTH is maximally suppressed, calcium absorption is greatest, bone loss rate reduced, and bone mineral density is highest.

Toxicity

The identified effects of vitamin D toxicity result from increased free circulating $1,25(OH)_2D$, which in turn causes increased absorption of dietary calcium and hypercalcaemia. Toxicity should not occur from endogenous over-production of vitamin D. This is because excess sunlight exposure, beyond what is required to produce maximal previtamin D_3, should result in biologically inert photoproducts being produced from previtamin D_3 and vitamin D_3 (3;9). Toxicity therefore usually only occurs from high intakes, mainly of supplements containing vitamin D. Long-term vitamin D supplementation of less than 25 µg per day has been shown to be non-toxic (9). Most studies which tested higher dosages of supplements have only been conducted for a short time period. One small intervention study of 12 patients administered dosages of vitamin D_3 which by the end of the 28 week study resulted in mean circulating 25(OH)D concentrations above 386 nmol/L (31). Vieth (32) suggests that adverse effects have not been reported with circulating 25(OH)D concentrations up to 140 nmol/L. Generally, major gaps exist in the current knowledge regarding the upper limits of toxicity. Most of the intervention studies testing this were small and lasted only for short time periods, consequently long-term and uncommon effects are unknown.

The picture is further complicated by results from recent prospective epidemiological studies where elevated all-cause mortality risks were observed at circulating 25(OH)D levels ranging from >93-140 nmol/L when compared against the mid-range reference categories (33-35).

Similarly, increased risks of overall cancer (35;36), prostate cancer (37-39) and pancreatic cancer (40) have also recently been reported in prospective studies when individuals within the highest circulating 25(OH)D levels were compared against those in reference categories with lower levels. Thus, intervention studies testing the long-term effects of supplement dosages over 25 µg are required.

1.1.5   Recommended vitamin D intakes

The recommended vitamin D intakes from the WHO/FAO and selected European countries are presented in Table 2. Current recommendations for vitamin D intakes are based on preventing rickets in children and osteomalacia in women of childbearing age.

In the UK, where latitudes range from 50-60°N, the majority of the country is unable to endogenously produce vitamin D_3 in winter months. Despite this, specific recommendations for men and non-childbearing women aged between 4 and 65 years are not in place. Generally, the different country recommendations are of a similar magnitude and have specific guidance for babies, children, pregnant and lactating women and the elderly.

Table 2. Recommendations for vitamin D intake from the WHO/FAO, United Kingdom, Nordic Countries and Spain

(Sources: SACN (30); Doets *et al,* (41); FAO/WHO (42)

WHO/FAO	Men (µg/day)	Women (µg/day)
0 months to 9 years	5	5
10 years to 50 years	5	5
51-65 years	10	10
65+ years	15	15
Pregnancy		5
Lactation		5
United Kingdom		
0-6 months	8.5	8.5
7 months to 3 years	7	7
4 years to 65 years	-	-
65+ years	10	10
Pregnancy	-	10
Lactation, 0-4 months	-	10
Lactation, 4+ months	-	10
Nordic countries		
9 months	10	10
5 years to 69 years	7.5	7.5
70+ years	10	10
Pregnancy	-	10
Lactation	-	10
Spain		
0 months to 9 years	10	
10 years to 49 years	5	
50 years	10	
70+ years	15	
Pregnancy		10
Lactation		10

1.2   Correlates and determinants of Vitamin D status

1.2.1   Skin synthesis

Endogenous production of vitamin D_3 requires sufficient skin exposure with the appropriate UVB radiation (wavelength 290-315 nm). Between mid-October and the start of April at latitudes of 52°N and above endogenous production of vitamin D_3 is not possible (43). This is because, in these months and at those locations, UVB does not reach the earth surface. The traditional belief was that the vitamin D endogenously produced in summer provides sufficient supplies for the winter months when production is not possible (in northern latitudes). Peak circulating 25(OH)D levels are usually observed in the late summer and early autumn months, while lowest levels are found at the end of winter months (9). An analysis of 7,437 Caucasian men and women from the 1958 British Birth Cohort reported lower mean circulating 25(OH)D levels for individuals who had their blood collected in the winter and spring months (41.1 nmol/L) compared to those who had blood collected in summer and autumn months (60.3 nmol/L) (27). Similarly, a southern Italian study reported lower mean 25(OH)D level for participants who had their blood collected in winter (42.7 nmol/L) compared to summer (84.0 nmol/L) (44). Finally, a small Norwegian cross-sectional analysis also reported lower mean 25(OH)D levels for participants who had their blood samples collected in March (51.5 nmol/L) compared to September (82.0 nmol/L) (45).

The current best estimates are that a fair skinned person at 40°N latitude requires 5-10 minutes sunlight exposure (during a sunny summers day) 2-3 times per week to achieve maximal previtamin D_3 production (9). For dark skinned individuals this time period increases to 30 minutes exposure.

Despite endogenous vitamin D_3 production not being possible above certain latitudes during periods of the year, using latitude as a predictor of vitamin D status seems only appropriate when comparing regions within certain European countries. For instance, in the UK, higher mean 25(OH)D levels were found among individuals from the southern regions of the country (42.6 nmol/L in winter/spring and 62.4 nmol/L in summer/autumn) compared to those from Scotland (35.4 nmol/L in winter/spring and 50.9 nmol/L in summer/autumn) (27). A similar latitude gradient has also been observed in a French study, where individuals located

in the south west of the country (94.0 nmol/L) had higher mean circulating 25(OH)D levels than those whose residence was in the north of the country (43.0 nmol/L) (46). However, when comparing circulating 25(OH)D levels across different European countries a reverse relationship has been observed. A recent systematic review of studies in Western, Northern and Southern Europe calculated a mean increase in circulating 25(OH)D of 11.8 nmol/L was associated with a 10 degree increase in latitude of country of residence (47).

Other factors which may affect vitamin D biosynthesis include clothing and sunscreen use. Covering of skin when exposed to sunlight can reduce the endogenous production of vitamin D_3. Hatun *et al.,* (48) observed that Turkish teenage girls who followed the Islamic dress code and wore a veil had a circulating 25(OH)D level half of that to those who were unveiled. Endogenous production of vitamin D is also reduced when sunscreen is used. The capacity of the skin to produce vitamin D_3 is lowered by up to 95% when sunscreen with sun protection factor (SPF) 15 is applied (3).

1.2.2   Skin pigmentation and ethnicity/race

Endogenous vitamin D production varies with ethnicity (9). The primary explanation for this is that dark skinned people have more melanin than light skinned people. Melanin acts as a natural sunscreen filtering out UVB radiation, meaning that darker skinned people need extended exposures in the sunlight, in comparison to light skinned people, to produce equivalent amounts of vitamin D_3. Another possible reason for differences in circulating 25(OH)D between ethnic groups is the cultural variation in the consumption of foods containing vitamin D.

1.2.3   Adiposity

An inverse association between higher body mass index (BMI), waist circumference, or obesity and circulating 25(OH)D concentrations has consistently been reported (27;49;50). The main biological reason for this is that vitamin D is fat soluble so therefore less bioavailable amongst obese people due to deposition in body fat compartments (51).

Another potential reason for this difference is that the decreased mobility associated with obesity causes less sunlight exposure and ultimately less endogenous vitamin D production.

1.2.4 Physical activity

Increased physical activity has been positively associated with increased circulating 25(OH)D (49;52). Whether this is because physical activity is a surrogate measure for sun exposure and a healthier lifestyle, or this is caused by a separate biological effect is not known.

1.2.5 Sex

Men have higher circulating 25(OH)D levels than women. Scragg *et al.,* (53) in an National Health and Nutrition Examination Survey III (NHANES III) analysis reported a mean circulating 25(OH)D of 78.8 nmol/L in men compared to 72.6 nmol/L in women. This difference in vitamin D status may partially be due to women having more body fat than men and circulating 25(OH)D being inversely correlated with adiposity. However, Scragg *et al.,* (53) adjusted for BMI and many of the other determinants highlighted in this section and the gender difference in circulating 25(OH)D remained.

1.2.6 Age

An inverse relationship between age and circulating 25(OH)D has been consistently observed. Amongst 6,228 participants of NHANES III, the mean circulating 25(OH)D for participants aged 20-39 years and over 60 years was 81 nmol/L and 69.5 nmol/L respectively (53). One possible explanation for this age related decline is that sunlight exposure and dietary intakes reduce as a person gets older. However, this does not explain

the observation that a 70 year old exposed to the same amount of sunlight as a 20 year old endogenously produces one quarter of vitamin D (3). This may be caused by an age dependent reduction in 7-dehydrocholesterol production in the epidermis of the skin where more than 80% of previtamin D_3 is produced (54). Another possible biological explanation is that renal production of $1,25(OH)_2D$ declines with age (9).

1.2.7   Smoking habits

The majority of previous research has reported smokers as having lower vitamin D status than non-smokers (49;50). Possible reasons for this difference include: a smoking induced increase in liver enzyme activity (55); a reduction in dermal production amongst smokers (9); and other differences in sun exposure, diet, and lifestyle factors which may exist between smokers and non-smokers.

1.2.8   Retinol

Vitamin A and vitamin D compete for the same receptor protein (retinoid X receptor). This means that high levels of vitamin A may inhibit vitamin D absorption (9). High intake of retinol has been associated with reduced bone mineral density, hip fracture increases, and increased fracture risk (56). Whether retinol has any deleterious effects on the anti-carcinogenic properties of vitamin D is unknown.

1.3 Previous epidemiological research: vitamin D and risks of cancer incidence and mortality

Epidemiological research has investigated the vitamin D and cancer association using differing approaches to vitamin D exposure assessment. Prospective studies that use circulating 25(OH)D measures are often viewed as the "gold standard" (9). This is because measurement of this metabolite encompasses endogenously produced vitamin D from UVB exposure and dietary/supplementary intakes. The overview of the relevant scientific literature below generally focuses on prospective nested case-control and cohort studies which have measured actual or predicted circulating 25(OH)D levels.

Note: some of the risk estimates from the previous epidemiological research have been inverted so higher levels of actual or predicted circulating 25(OH)D are consistently compared against lower levels.

1.3.1 Overall cancer

Incidence

Few studies have examined the association between vitamin D and risk of total cancer incidence and the results are inconsistent. The mixed findings may be due to the heterogeneous relationships between vitamin D and the incidences of the individual cancer types. Using a model of predicted circulating 25(OH)D in the Health Professionals Follow-Up Study (HPFS), Giovannucci *et al.,* (52) reported a 16% lower risk (95% CI: 0.72-0.98) of total cancer incidence associated with a 25 nmol/L increment. A recent German cohort analysis also observed a reduced overall cancer incidence risk amongst men in the highest quartile of circulating 25(OH)D when compared against those in the lowest group (HR 0.75, 95% CI: 0.60-0.94); although, amongst women, no association was observed (57). In a small Swedish cohort, a U-shaped relationship for total cancer incidence was observed (as was for cancer mortality and all-cause mortality) (35). In the fully adjusted model a 68% increased risk of cancer incidence was observed amongst those in the highest circulating 25(OH)D category when compared against the mid-range reference category (>98 vs. 46-93 nmol/L, RR 1.68, 95% CI: 1.06-2.65). Increased risk was also observed amongst those in the lowest 25(OH)D category (<39 vs. 46-93 nmol/L, RR 1.65, 95% CI: 1.08-2.54) (35).

One small randomised control trial (RCT) has been conducted where women in the intervention group received 27.5 µg of vitamin D_3 plus 1.5 g of calcium daily over 4 years (58). A suggestive, but statistically non-significant, inverse association was observed for vitamin D and risk of cancer incidence (RR 0.59, 95% CI: 0.32-1.10). The association was statistically significant when the intention to treat analysis was conducted (*P*-value <0.03). However, the trial methodology has been criticised and results are inconclusive (9).

Mortality

The limited studies investigating the association between circulating 25(OH)D and overall cancer mortality have produced variable results. Especially of interest is the unexpected elevated risk of death at high circulating 25(OH)D levels. In an NHANES III analysis, Freedman *et al.,* (36) reported an increased risk amongst men with circulating 25(OH)D levels above 100 nmol/L when compared versus those with the lowest concentrations (<37.5 nmol/L) (RR 1.85, 95% CI: 1.02-3.35; *P*-trend 0.09); although, no such association was observed amongst women (RR 0.64, 95% CI: 0.35-1.18; *P*-trend 0.29). One smaller U.S. study reported a near significance 92% increased risk (HR 1.92, 95% CI: 1.00-3.70; *P*-trend 0.086) of cancer death amongst men in the highest circulating 25(OH)D exposure category when compared against those in the lowest category (>74.9 vs. <49.7 nmol/L) (54). Finally, in a small Swedish cohort of elderly men, an increased risk was observed amongst those with the highest levels of circulating 25(OH)D (>93 nmol/L) when compared against those in the mid-range reference category (46-93 nmol/L) (HR 2.45, 95% CI: 1.36-4.00) (35). A U-shaped relationship was observed in this study, as a risk estimate of similar magnitude was observed when participants in the lowest exposure category (<39 nmol/L) were compared against those in the reference group (HR 2.34, 95% CI: 1.36-4.00) (35).

In contrast, several studies have reported that higher levels of actual circulating, or predicted 25(OH)D, are associated with lower cancer mortality risk. First, in the HPFS, a 29% lower cancer mortality risk (HR 0.71, 95% CI: 0.60-0.83) was associated with a 25 nmol/L increment in predicted circulating 25(OH)D (52). Most recently, a German prospective analysis observed a 30% decreased cancer mortality risk (HR 0.70, 95% CI: 0.53-0.93) when those with circulating 25(OH)D levels above 50 nmol/L were compared against those with levels below 30 nmol/L (59). Another small cohort analysis reported a significantly reduced risk amongst men and women in the highest 25(OH)D exposure category (>57.5 vs. <25 nmol/L, HR 0.45, 95% CI: 0.22-0.93) (60). Finally, an analysis of older men within the

Whitehall II cohort, reported a near-significance inverse association between circulating 25(OH)D and cancer mortality risk (61).

The remaining five prospective studies published to date have reported statistically non-significant associations (62-65). The largest of these studies was a recent analysis in the U.S. based Southern Community Cohort Study (Q4 vs. Q1, OR 0.78, 95% CI: 0.47-1.28; *P*-trend 0.53) (64). Similarly, in the WHI observational study, a weak non-significant decreased cancer mortality risk was observed amongst the post-menopausal women (Q4 vs. Q1, HR 0.72, 95% CI: 0.46-1.14; *P*-trend 0.11) (65).

Overall, mixed results have been observed in the prospective studies to date and the role of vitamin D on cancer mortality is uncertain. Further studies are required to investigate this relationship.

1.3.2   Colorectal cancer

Incidence

The vitamin D-cancer hypothesis was first proposed from the results of an ecological study which reported positive associations between latitude and colon cancer mortality (1). Since then, numerous studies for colorectal cancer incidence and mortality have been published which have used dietary vitamin D or measured circulating 25(OH)D as exposure measures. The totality of the epidemiological research suggests that high vitamin D status may lower the risk of colorectal cancer incidence.

The majority of dietary studies have reported statistically non-significant reductions in risk. For instance, in the E3N cohort of women a non-significant 11% reduced risk (HR 0.89, 95% CI: 0.58-1.36) was observed amongst participants with the highest daily intake (>3.23 µg) compared with those in the lowest intake category (<1.72 µg) (66). Also, a recent EPIC nested case-control study reported non-significant reduced risk amongst participants with the highest dietary intakes (≥5.8 vs. <2.1 µg per day, RR 0.84, 95% CI: 0.60-1.17; *P*-trend

0.19) (67). Two cohort studies have reported statistically significant associations amongst men but not women. Firstly, in the Cancer Prevention Study II Nutrition Cohort (United States) (CPS-II), a reduced risk amongst men in the highest intake quintile (>13.1 µg per day) was observed when compared against those in the lowest intake quintile (<2.8 µg per day) (RR 0.71, 95% CI: 0.51-0.98; *P*-trend 0.02) (68); no association was observed for women. Similarly, Park *et al.,* (69), in a Multiethnic Cohort Study analysis reported a 28% (HR 0.72, 95% CI: 0.51-1.00; *P*-trend 0.03) reduced risk of colorectal cancer amongst men in the highest intake quintile (>6.9 µg per day) versus those in the lowest (<1 µg per day); once more, no association observed amongst women.

The largest prospective study which measured circulating 25(OH)D was a nested case-control study using EPIC data (67). In this analysis circulating 25(OH)D was categorised in five pre-defined clinically relevant groups (<25; ≥25 to <50; ≥50 to <75; ≥75 to <100; ≥100 nmol/L). When colorectal cancer risk was analysed, the estimates did not reach statistical significance but were suggestive of elevated risk at lower circulating 25(OH)D levels, and reduced risk at higher circulating levels (*P*-trend <0.001). However, when colon and rectal cancer were analysed separately divergent associations were observed. For colon cancer, a statistically significant increased risk was observed at the lowest circulating 25(OH)D levels (<25 vs. ≥50 to <75 nmol/L, OR 1.90, 95% CI: 1.10-3.29; *P*-trend <0.001), and a non-significant reduced risk was observed at the highest level (≥100 vs. ≥50 to <75 nmol/L, OR 0.71, 95% CI 0.46-1.08); in contrast, no association was observed for rectal cancer (67). A similar relationship for colon cancer was also reported in a pooled analysis of the Nurses' Health Study (NHS) and HPFS data (70). When participants in the highest circulating 25(OH)D group were compared against those in the lowest, a 54% statistically significant reduced risk was observed (OR 0.46, 95% CI: 0.24-0.89; *P*-trend 0.005). Similarly to the EPIC study, no associations were reported for rectal cancer (70).

A Multiethnic cohort nested case-control study observed an inverse association for colorectal cancer (OR per doubling of circulating 25(OH)D, 0.68; 95% CI: 0.51-0.92) (71). Whereas two other nested case-control studies reported non-significant inverse risk estimates for colon cancer (72;73). However, no association for colorectal cancer was observed in the Physicians' Health Study (PHS) (Q4 vs. Q1, OR 1.08, 95% CI: 0.62-1.87) (74). Whereas a recent Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC) nested case-control study did not observe lower colorectal cancer risk for higher 25(OH)D levels; instead a suggestive elevated risk at lower levels was observed (<25 vs. 50-75 nmol/L, OR 0.68,

95% CI: 0.45-1.03) (75). Nor was an association observed in a small Japanese nested case-control study (76). However, overall, in a recent meta-analysis, a 5.9 nmol/L increment in circulating 25(OH)D was associated with a 4% lower risk (95% CI: 0.94-0.97) of colorectal cancer (77); similar risk estimates were also yielded for colon and rectal cancers.

To date, two RCTs have examined the influence of vitamin D supplementation on colorectal cancer risk (78;79). The largest was the Women's Health Initiative (WHI) clinical trial which included 36,282 post-menopausal women (79). During the 7 year duration of the trial, the intervention group were given 10 µg of vitamin D plus 1 g of elemental calcium each day. By the end of the trial 168 and 154 colorectal cancer cases were reported in the intervention and placebo groups respectively, and the final results were null (HR 1.08, 95% CI: 0.86-1.34; *P*-value 0.51). Criticisms of the trial methodology have been used to explain this null result. First, the duration of 7 years may be too short a time period for vitamin D to influence colorectal cancer occurrence (80). Second, the dosage given to intervention subjects may have been too low to differentiate them from the placebo control group (9). Finally, the contrast in circulating 25(OH)D levels between the intervention group and the placebo group was lower than expected due to a high proportion of women taking non-study supplements (80). A small UK RCT which administered vitamin D supplements to examine the prevention of osteoporotic fractures, also, in secondary analyses, recorded cases of colon cancer amongst subjects (78). The intervention group of the trial was administered with the daily equivalent of 41 µg per day over a 5 year period. Once more no reduction in colon cancer was observed amongst the intervention group (HR 1.02, 95% CI: 0.60-1.74; *P*-value 0.94).

Overall, the observational evidence suggests that vitamin D may have a protective colorectal cancer role. However, these observations have yet to be confirmed with intervention studies.

Mortality

Few studies have investigated the relationship between vitamin D and colorectal cancer mortality; although the limited available evidence indicates that vitamin D may have an important role in limiting tumour progression. Some studies suggest a greater reduction of risk at higher exposures to vitamin D for mortality than incidence (80). A U.S. ecological study reported higher risk estimates for male colon cancer mortality (RR 1.27 95% CI: 1.24-1.30) than incidence (RR 1.11, 95% CI: 1.08-1.13) when the northern regions of the country

were compared against the southern (81); similar associations were observed for women, and also for rectal cancer in both sexes.

Most prospective studies have measured pre-diagnostic circulating 25(OH)D to assess the relationship with colorectal cancer mortality. In an NHANES III analysis, a non-significant reduced colorectal cancer mortality risk was observed when participants in the highest and lowest exposure groups were compared (≥100 vs. <50 nmol/L, RR 0.35, 95% CI: 0.11-1.14; *P*-trend 0.09) (36). Ng *et al.,* (82), in a pooled NHS and HPFS analysis of 304 participants with colorectal cancer, reported a non-significant reduced risk of death for participants in the highest 25(OH)D quartile when compared against the lowest (Q4 vs. Q1, HR 0.61, 95% CI: 0.31-1.19; *P*-trend 0.23). Within EPIC, pre-diagnostic circulating 25(OH)D levels amongst participants who developed colorectal cancer were also associated with a subsequent reduced mortality risk from the disease (n=444 deaths; Q5 vs. Q1, HR 0.69, 95% CI: 0.50-0.93; *P*-trend 0.04) (83).

To date, one small Japanese study has measured post-diagnosis levels of actual circulating 25(OH)D to assess the relationship (84); amongst the 257 colorectal cancer patients, higher levels of 25(OH)D were associated with improved survival (HR per 2.5 nmol/L increment 0.91, 95% CI: 0.84-0.99). Similarly, a NHS and HPFS analysis, which predicted post-diagnosis levels of circulating 25(OH)D levels for 1,017 colorectal cancer patients reported a 50% reduced risk (HR 0.50, 95% CI: 0.26-0.95; *P*-trend 0.02) for participants within the highest category (>77 nmol/L) when compared against those in the lowest category (<64 nmol/L) (85).

Overall, the limited data does suggest that vitamin D levels are inversely associated with colorectal cancer mortality. However, the few previous studies have been small and larger analyses are required.

1.3.3   Prostate cancer

Incidence

The large body of evidence investigating the association between vitamin D and prostate cancer does not support the hypothesis that high vitamin D exposures can reduce incidence. Ecological studies first highlighted an inverse association between UVB radiation and prostate cancer mortality (86). Although this association was not supported by dietary intake studies, where one cohort (9.2 vs. 2.4 µg, RR 0.80, 95% CI: 0.50-1.30; *P*-trend 0.86) (87) and two case-control studies (88;89) failed to observe any associations.

The vast majority of prostate cancer nested case-control studies which measured circulating 25(OH)D have yielded null or non-significant positive associations (90-99). In a previous EPIC analysis, a suggestive increased risk was observed when the highest versus lowest quintiles of circulating 25(OH)D were compared (OR 1.28, 95% CI: 0.88-1.88; *P*-trend 0.19) (97). Significant positive prostate cancer associations have been observed in three Nordic nested case-control studies. Firstly, Tuohimaa *et al.,* (37), reported an elevated risk of circulating 25(OH)D levels greater than 80 nmol/L, when compared against the mid-range (40-59 nmol/L) category (OR 1.70, 95% CI: 1.10-2.40); at low concentrations (≤19 nmol/L), increased risk was also reported (OR 1.50, 95% CI: 0.80-2.70), although this association was not statistically significant. Secondly, in an ATBC study analysis, a 56% increased prostate cancer risk (OR 1.56, 95% CI: 1.15-2.12; *P*-trend 0.01) was observed when the highest and lowest 25(OH)D exposure groups were compared (38). Finally, in the largest prospective study to date (n=2,106 cases) which measured circulating 25(OH)D, a 30 nmol/L higher increment was associated with a 13% greater (95% CI: 1.02-1.25) prostate cancer risk (39). Overall, a recent meta-analysis, which did not include the latter two of these studies with positive associations, hinted at a possible weak positive relationship, as a 25 nmo/L increment in circulating 25(OH)D was associated with an OR of 1.04 (95% CI: 0.99-1.10) (100). Overall, the evidence is unsupportive of a possible protective role for vitamin D on prostate cancer; instead null or non-significant positive associations have usually been reported.

Mortality

The effect of vitamin D on prostate cancer mortality has rarely been studied. An NHANES III analysis (74 prostate cancer deaths) which used pre-diagnosis circulating 25(OH)D measurements did not find an association (80-<100 vs. <50 nmol/L, RR 1.23, 95% CI: 0.50-3.05; *P*-trend 0.84) (36). However, a small Norwegian cohort study (52 deaths) which collected post-diagnosis circulating 25(OH)D measurements from patients, reported a 67% significantly reduced risk (95% CI: 0.14-0.77) of prostate cancer mortality when the medium and lowest exposure categories were compared (50-80 vs. <50 nmol/L) (101). Finally, a more recent analysis in the HPFS study (114 deaths) used pre-diagnostically measured 25(OH)D to assess the relationship with lethal prostate cancer and reported inverse linear associations (Q4 vs. Q1, OR 0.43, 95% CI:0.24-0.76; *P*-trend 0.001) (102). Overall, few studies have examined vitamin D with prostate cancer mortality end-points and those that have been conducted have been small; that said, the results indicate that a possible inverse relationship may be present, although larger studies are required to confirm this association.

1.3.4   Breast cancer

Incidence

Ecological studies have reported inverse associations between UVB and breast cancer incidence and mortality (103;104). However, mixed results have been observed in observational studies. Several case-control studies have reported a significantly reduced breast cancer risk at higher circulating 25(OH)D levels (105-109). In a U.S. study of pre- and post-menopausal women, a 44% reduced risk (OR 0.56, 95% CI: 0.41-0.78; *P*-trend 0.004) was observed when participants in the highest exposure category were compared against those in the lowest (109). A German study in post-menopausal women reported a significantly reduced risk when the highest circulating 25(OH)D category was compared versus the lowest (≥75 vs. <30 nmol/L, OR 0.31, 95% CI: 0.24-0.42; *P*-trend <0.0001) (108).

In contrast to these inverse associations, nested case-control studies have generally not reported associations, except for two recent smaller studies (110;111). Firstly, a small Danish study (n=120 cases) observed reduced risk amongst pre-menopausal women when

participants in the highest and lowest tertiles (>84 vs. <60 nmol/L) of circulating 25(OH)D were compared (OR 0.38, 95% CI: 0.15-0.97) (110). Secondly, a E3N cohort analysis observed a 27% reduced risk (95% CI: 0.55-0.96; *P*-trend 0.02) amongst participants in the highest tertile of 25(OH)D when compared against the lowest tertile (67.4 vs. <49.4 nmol/L) (111). However, these inverse significant associations have not been replicated in larger nested case-control studies (112-117). For instance, a WHI analysis, which included 1,067 cases, did not report an association (Q5 vs. Q1, OR 1.22, 95% CI: 0.89-1.67; *P*-trend 0.20) (114). Furthermore, in an EPIC analysis, which was the largest study to date (n=1,391 cases), no association was observed when the highest and lowest quintiles of circulating 25(OH)D were compared (>63 vs. ≤39.3 nmol/L, OR 1.07, 95% CI: 0.85-1.36; *P*-trend 0.67) (117). Despite these largely null results of individual nested case-control studies, a recent meta-analysis - published prior to the EPIC analysis – reported a pooled OR estimate of 0.87 (95% CI: 0.77-0.99) for the highest versus the lowest circulating 25(OH)D quintiles (118).

One vitamin D-breast cancer RCT has been conducted. The WHI clinical trial also recorded breast cancer incidence amongst subjects. Similarly to colorectal cancer, no reduction in breast cancer incidence was observed at the end of the 7 year intervention (HR 0.96, 95% CI: 0.85-1.09) (114).

Generally, the results of case-control studies have largely been consistent and indicate a possible protective role for vitamin D; whilst the mixed results from nested case-control studies require further investigation in larger prospective analyses.

Mortality
One study which measured pre-diagnosis circulating 25(OH)D levels has been conducted. An NHANES III analysis of pre- and post-menopausal women observed a non-significant reduced breast cancer mortality risk when individuals with circulating 25(OH)D of 80-<100 nmol/L were compared against those within the lowest reference category (vs. <50 nmol/L, RR 0.65, 95% CI: 0.18-2.38) (36). Other studies have measured circulating 25(OH)D levels after diagnosis. Goodwin *et al.,* (119) reported a near-significance HR of 0.63 (95% CI: 0.38-1.04) when patients in the highest and lowest circulating 25(OH)D categories (>72 vs. <50 nmol/L) were compared against each other. Similarly, a near significance reduced mortality risk was reported amongst German post-menopausal women with high circulating 25(OH)D

levels when compared against the mid-range reference category (≥55 vs. <35 nmol/L, HR 0.65, 95% CI: 0.42-1.00; *P*-trend 0.07) (120). However, a beneficial prognostic role for vitamin D post breast cancer diagnosis was not observed in the Women's Healthy Eating and Living (WHEL) Study, where no association was observed between circulating 25(OH)D levels and recurrence of the disease (≥75 vs. 25 nmol/L, OR 0.88, 95% CI: 0.43-1.75) (121).

Overall, whether vitamin D has a role on breast cancer mortality is uncertain as few previous studies have been conducted, and those that have been published have generally been small.

1.3.5   Pancreatic cancer

The prospective data investigating the relationship between circulating 25(OH)D and pancreatic cancer risk are inconsistent. An ATBC study of 200 men who smoked, reported an OR of 2.92 (95% CI: 1.56-5.48; *P*-trend 0.001) for participants in the highest circulating 25(OH)D quintile when compared against those in the lowest quintile (122). This significant positive association was not replicated in the United States (U.S.) based Prostate, Lung, Colorectal, and Ovarian Screening Trial (PLCO) study (Q5 vs. Q1, OR 1.45, 95% CI: 0.66-3.15; *P*-trend 0.49) (123). It was suggested that this inconsistency was due to the Finnish ATBC participants being smokers (a major pancreatic cancer risk factor) and living at higher latitudes than participants in the U.S. PLCO study (124).

The recent Vitamin D Pooling Project of Rarer Cancers (VDPP) included data from eight worldwide prospective cohorts (including the ATBC and PLCO studies). A nested case-control study from the pooling project used clinically defined circulating 25(OH)D exposure categories within their models (40). When the highest and mid-range clinically defined exposure categories (≥100 vs. 50-<75 nmol/L) were compared, elevated pancreatic cancer risk was observed once more (OR 2.12, 95% CI: 1.23-3.64). No association was found for the lowest exposure group when compared against the mid-range (<25 vs. 50-<75 nmol/L, OR 0.95, 95% CI: 0.68-1.32). Importantly, when the ATBC cohort was excluded in sensitivity analyses, a similar association was observed in the highest circulating 25(OH)D category

(≥100 nmol/L) (OR 2.95, 95% CI: 1.47-5.93). Also, when limited to U.S. only cohorts, the association remained in the highest exposure group (OR 2.98, 95% CI: 1.48-6.02) (40). However, other recent U.S. based prospective analyses have yielded inverse associations (124;125). Firstly, predicted circulating 25(OH)D was inversely associated with pancreatic cancer risk in the full NHS and HPFS cohorts (RR 0.65, 95% CI: 0.50-0.86; *P*-trend 0.001) (124). More recently, a pooled analysis of five U.S. based nested case-control studies (HPFS, NHS, PHS, WHI, and the Women's Health Study (WHS)) which measured circulating 25(OH)D, reported an inverse association when individuals in the highest and lowest quintiles were compared (>81 vs. <45 nmol/L, OR 0.67, 95% CI: 0.46-0.97; *P*-trend 0.03) (125). In the same analysis, when identical clinically defined exposure categories used by the VDPP were used, elevated risk was not observed at 25(OH)D levels above 100 nmol/L when compared against the mid-range reference category (vs. 50-<75 nmol/L, OR 1.01, 95% CI: 0.63-1.62).

Overall, the data reveal a mixed and confusing picture for the vitamin D-pancreatic cancer relationship, underlying the importance of additional studies.

1.3.6 Lung cancer

Null lung cancer results have been observed in the three previous prospective studies which measured circulating 25(OH)D. In a small Finnish cohort (n=122 cases), no association was observed for lung cancer incidence when the highest and lowest tertile of circulating 25(OH)D were compared (RR 0.72, 95% CI: 0.43-1.19; *P*-trend 0.22) (126). Similarly, in the ATBC study (n=500 cases), no association was reported (Q5 vs. Q1, OR 1.08, 95% CI: 0.67-1.75) (127). Finally, an NHANES III analysis observed no association for lung cancer mortality (n=258 deaths) when the highest and lowest quartiles of circulating 25(OH)D were compared (≥80.3 vs. <44 nmol/L, RR 0.95, 95% CI: 0.62-1.44) (128). The sparse data mean that whether vitamin D has any lung cancer role is unknown and further prospective studies are required. Importantly, these larger analyses should (where sample size allows) assess the vitamin D-lung cancer relationship amongst never smokers, as risk estimates for lung cancer may be biased by smoking habits, or alternatively whether any benefit may be greater to smokers.

1.3.7 Kidney cancer

A VDPP nested case-control study (n=775 cases) reported null results when individuals in the low (<25 vs. 50-<75 nmol/L, OR 0.94, 95% CI: 0.64-1.37) and high (≥100 vs. 50-<75 nmol/L, OR 0.92, 95% CI: 0.44-1.92) circulating 25(OH)D groups were compared against those in the mid-range reference group (*P*-trend 0.86) (129). However, an analysis of the NHS and HPFS cohorts (n=408 cases), recently reported that a 25 nmol/L increment in predicted circulating 25(OH)D was associated with a 44% lower kidney cancer risk (95% CI: 0.42-0.74) (130).

1.3.8 Stomach and oesophageal cancers

Few studies have analysed the relationships between circulating 25(OH)D and stomach and oesophageal cancers. An VDDP nested case-control study reported null results when individuals with low levels (<25 vs. 50-<75 nmol/L, OR 0.90, 95% CI: 0.65-1.24) and high levels (≥100 vs. 50-<75 nmol/L, OR 0.81, 95% CI: 0.39-1.69) of circulating 25(OH)D were compared against the mid-range reference category (131). Similar null results were also observed when oesophageal and gastric cancer were analysed separately (131). In contrast, a Chinese nested case-control study reported a statistically significant 77% (OR 1.77, 95% CI: 1.16-2.70; *P*-trend 0.003) increased oesophageal squamous cell carcinoma risk amongst men when the highest circulating 25(OH)D quartile was compared against the lowest (132); although no associations were observed for other oesophageal cancer disease subtypes. Finally, in the HPFS cohort, higher predicted circulating 25(OH)D was associated with significantly lower risk of oesophageal cancer (HR per 25 nmol/L increment in predicted 25(OH)D, 0.37, 95% CI: 0.17-0.80) and non-significantly lower risk of stomach cancer (HR per 25 nmol/L increment in predicted 25(OH)D, 0.58, 95% CI: 0.26-1.33) (52).

1.3.9 Bladder cancer

To date, two prospective nested case-control studies have been conducted with divergent results reported within them. Firstly, a significant linear inverse association was observed in an ATBC cohort study (≥50 vs. <25 nmol/L OR 0.58, 95% CI: 0.34-0.98; *P*-trend 0.04) (133).

In contrast, an analysis within the PLCO cohort reported no association between circulating 25(OH)D and bladder cancer risk (Q4 vs. Q1, OR 1.20, 95% CI:0.74-1.92; *P*-trend 0.56) (134). No association was also observed in the HPFS cohort between predicted 25(OH)D and bladder cancer (52).

1.4.0   Non-Hodgkin's lymphoma

A VDPP nested case-control analysis, reported null non-Hodgkin lymphoma (NHL) associations for participants in the lowest (<25 vs. 50-<75 nmol/L, OR 1.08, 95% CI: 0.78-1.50) and highest (≥100 vs. 50-<75 nmol/L, OR 0.86, 95% CI: 0.57-1.27) circulating 25(OH)D category when compared against the mid-range reference group (*P*-trend 0.68) (135). A null result was also reported within the HPFS study when predicted 25(OH)D was used to assess vitamin D status (52). While a recent EPIC nested case-control study of 1,127 cases also observed a null result for overall lymphoid cancers (Q4 vs. Q1, OR 1.05, 95% CI: 0.81-1.38; *P*-trend 0.52) (136). However, for B-NHL, the most common disease subtype, a significant positive association was observed within EPIC (≥75 vs. 50-<75 nmol/L, OR 1.36, 95% CI: 1.00-1.83; *P*-trend 0.007) (136).

1.4.1   Skin cancer

For overall skin cancer incidence, a near significance positive association was observed for predicted circulating 25(OH)D in the HPFS study (52). Other studies have stratified skin cancer by disease subtype. For melanoma, non-significant positive associations have been observed in two previous nested case-control studies (137;138), the largest of which was an ATBC analysis (≥50 vs. <25 nmol/L, OR 1.32, 95% CI: 0.64-2.72; *P*-trend 0.51) (137). For the relationship between 25(OH)D and overall non-melanoma skin cancer (NMSC), a small study (178 cases) of elderly men reported an inverse association (Q5 vs. Q1, OR 0.54, 95% CI: 0.31-0.96; *P*-trend 0.04) (139). A NHS analysis reported positive associations for both basal cell carcinoma (BCC; <75 vs. >50 nmol/L, OR 2.07, 95% CI: 1.58-2.80; *P*-trend <0.0001) and squamous cell carcinoma (SCC; <75 vs. >50 nmol/L, OR 3.77, 95% CI: 1.70-8.36; *P*-trend 0.0002) when the highest and lowest quartiles of circulating 25(OH)D were

compared (140). However, a smaller Australian study reported a positive association for BCC (≥75 vs. <75 nmol/L, OR 1.51, 95% CI: 1.10-2.07), but not SCC (≥75 vs. <75 nmol/L, OR 0.67, 95% CI: 0.44-1.03) (138). The contrasting results between studies may be influenced by the vitamin D-skin cancer relationship being confounded by sun exposure habits/behaviours; the studies highlighted above adjusted for sun exposure covariates to varying degrees.

1.4.2   Ovarian cancer and endometrial cancer

Four previous nested case-control studies have analysed the relationship between circulating 25(OH)D and ovarian cancer risk (141-143). The largest was a VDPP analysis which included 516 cases and reported non-significant associations at low (<25 vs. 50-<75 nmol/L, OR 1.08, 95% CI: 0.64-1.81) and high (≥100 vs. 50-<75 nmol/L, OR 1.11, 95% CI: 0.61-2.05) circulating 25(OH)D levels when compared against the mid-range reference category (*P*-trend 0.65) (143). Similarly, the other nested case-control studies did not observe any associations (141;142). A recent meta-analysis of prospective studies also reported a non-significant risk estimate (RR per 50 nmol/L of 25(OH)D, 0.83, 95% CI: 0.63-1.08) (144).

Data for vitamin D and endometrial cancer are sparse, with only one nested case-control study being conducted to date. In this VDPP analysis, non-significant associations were reported when individuals in the low (<25 vs. 50-<75 nmol/L, OR 1.02, 95% CI: 0.68-1.53) and high (≥100 vs. 50-<75 nmol/L, OR 0.85, 95% CI: 0.47-1.53) circulating 25(OH)D groups were compared against those in the mid-range reference group (*P*-trend 0.81) (145). A null result was also reported in an NHS study which used predicted 25(OH)D levels to assess status (Q5 vs. Q1, HR 1.00, 95% CI: 0.73-1.36) (146).

1.4.3   Other less common cancers

Prospective data on the relationships between vitamin D and other less common cancers in Western populations are absent. For instance, no previous nested case-control or cohort

analyses have been carried out for liver and thyroid cancers; whilst for brain cancer, the only previous analysis was in the HPFS, where a non-significant positive association observed with predicted circulating 25(OH)D (52).

1.5 Previous epidemiological research: vitamin D and risks of all-cause and cause-specific mortality

1.5.1 All-cause mortality

The available prospective data suggests that a low vitamin D status may contribute to elevated all-cause mortality risk. In studies which solely investigated the circulating 25(OH)D-all-cause mortality relationship using linear models, consistent inverse associations have been observed when the high and low exposure groups are compared. One of the largest prospective studies to date, in terms of the number of all-cause deaths (n=3,215 deaths), was of elderly men (mean baseline age of 77 years) who were participants of the Whitehall II cohort (61). In this analysis the all-cause mortality HR for a twofold higher measure of circulating 25(OH)D was 0.78 (95% CI: 0.72-0.85) (61). Similarly, the Southern Community Cohort Study nested case-control study, which included 1,852 all-cause deaths, observed a reduced all-cause mortality risk when the highest quartile of circulating 25(OH)D was compared against the lowest quartile (OR 0.56, 95% CI: 0.44-0.70) (64); with similar associations observed amongst African American and non-African American participants. Another larger cohort study (n=1,083 deaths) reported a 40% reduced all-cause mortality risk (HR 0.60, 95% CI: 0.50-0.71) when individuals with high and low concentrations of circulating 25(OH)D were compared (>50 vs. <30 nmol/L) (59).

Other smaller prospective studies have observed similar inverse linear associations (63;147-151); although a number of other small studies have reported non-significant associations (54;65;152). For instance, in a WHI observational study analysis, a weak non-significant reduced all-cause mortality risk was observed amongst the post-menopausal women participants (Q4 vs. Q1, HR 0.80, 95% CI: 0.51-1.25; *P*-trend 0.39) (65).

Meta-analyses for all-cause mortality have yielded similar risk estimates despite differing study inclusion criteria. A 28% reduced all-cause mortality risk was observed when the highest and lowest quartiles were compared (61). Whereas Schöttker *et al.,* (59), in a linear dose-response analysis, reported a HR of 0.92 (95% CI: 0.89-0.95) being associated with a 20 nmol/L higher level of circulating 25(OH)D.

However, three studies with a larger number of deaths and wider ranges of circulating 25(OH)D levels, have observed a J- or U-shaped relationship, with greater all-cause mortality risk at higher as well as lower 25(OH)D levels (33-35). The largest of these studies was a Danish retrospective cohort study which included 247,574 participants amongst whom 15,198 deaths were recorded (33). Within this study, a J-shaped risk curve was observed, with a 42% increased all-cause death risk (RR 1.42, 95% CI: 1.31-1.53) found amongst participants with circulating 25(OH)D above 140 nmol/L when compared against the 50 nmol/L reference category (33). While in an NHANES III analysis of the general population, a U-shaped risk relationship was present, with a non-significant increased all-cause mortality risk observed at circulating levels 25(OH)D above 125 nmol/L when compared against the mid-range reference category (75-97.5 nmol/L) (34). Finally, a U-shaped relationship and increased risk at higher circulating 25(OH)D was also observed in a Swedish cohort of elderly men (35); amongst those with the highest levels of circulating 25(OH)D (>93 nmol/L), an increased risk was also observed when compared versus the reference category (vs. 46-93 nmol/L, HR 1.57, 95% CI: 1.12-2.19) (35).

A meta-analysis of smaller RCT's in which a supplementary vitamin D intervention was administered, reported a significantly reduced occurrence of all-cause mortality (153). Amongst the 57,311 participants, 4,777 all-cause deaths occurred across the 18 studies, and a summary RR of 0.93 (95% CI: 0.87-0.99) was reported for participants who received vitamin D supplements.

Generally, in smaller studies, reduced all-cause death risks have been consistently observed with higher levels of circulating 25(OH)D, and this observational evidence has been further supported by the aforementioned meta-analysis of small RCT's. However, three larger observational studies have observed J- or U-shaped relationships, with increased all-cause

mortality risks associated with higher 25(OH)D levels. This possible detrimental effect of higher levels of vitamin D requires further investigation in other large prospective studies.

1.5.2   Circulatory disease mortality

The relationships between circulating 25(OH)D and circulatory/cardiovascular disease mortality have been studied frequently in prospective studies. In the largest study to date (n=1,358 vascular deaths) - an analysis of older men in the Whitehall II cohort - the vascular mortality HR for a twofold higher measure of circulating 25(OH)D was 0.80 (95% CI: 0.70-0.91); with similar risk estimates observed for ischaemic heart disease, strokes and other vascular deaths, when analysed separately (61). A similar inverse association was also observed in a Finnish cohort analysis in which 933 cardiovascular disease deaths were recorded amongst the men and women participants during the follow-up period (Q5 vs. Q1, HR 0.76, 95% CI: 0.61-0.95; *P*-trend 0.005) (154). In the NHANES III study, a reduced cardiovascular disease mortality risk was observed when the highest and lowest quartiles of circulating 25(OH)D were compared (>80 vs. <45 nmol/L. HR 0.78, 95% CI: 0.68-0.90) (34). Other smaller studies have also observed significant inverse associations between circulating 25(OH)D and circulatory disease mortality end-points (64;155;156). In the U.S. based Southern Community Cohort Study (n=531 circulatory disease deaths), reduced circulatory disease mortality risks were reported amongst African American (OR 0.63, 95% CI: 0.47-0.83; *P*-trend 0.003) and non-African American (OR 0.47, 95% CI: 0.31-0.72; *P*-trend <0.001) participants in the highest quartile of circulating 25(OH)D when compared against individuals in the lowest quartile (64).

Non-significant associations have also been observed in a number of smaller studies (35;54;59;63;65;157). Most recently, a prospective German analysis (n=350 cardiovascular disease deaths) reported a non-significant reduced cardiovascular disease mortality when participants with the highest and lowest levels of circulating 25(OH)D were compared (>50 vs. <30 nmol/L, HR 0.78, 95% CI: 0.57-1.06) (59). However, a recent meta-analysis of 18 prospective studies, many of which are detailed above, reported a 21% reduced overall vascular mortality risk (RR 0.79, 95% CI: 0.72-0.87) when the highest and lowest quartiles were compared (61).

Overall, the prospective evidence is indicative of an inverse relationship between circulating 25(OH)D and circulatory disease deaths. However, the studies conducted to date have been relatively small, with only two having more than 1,000 circulatory disease deaths recorded during the follow-up period. Thus, larger studies are warranted as the association can then be analysed across subgroups of other circulatory disease risk factors which many confound (e.g. smoking habits, BMI and age) the relationship.

1.5.3   Digestive disease mortality

The relationship between vitamin D and digestive disease mortality has seldom been studied. One small Danish cohort analysis, with just 34 digestive disease deaths, reported an inverse association with circulating 25(OH)D (Q4 vs. Q1, HR 0.28, 95% CI: 0.10-0.78; *P*-trend 0.004) (62). For liver disease mortality, a non-significant inverse association was also observed with circulating 25(OH)D among 75 patients with chronic liver failure (158). Other small studies have also reported vitamin D deficiency amongst liver disease patients (159-161), but whether this is a causal association or a consequence of the disease – due to compromised digestive and liver functions – is unclear. Larger studies, with sufficient cases and longer follow-up times are required to further scrutinise this relationship.

1.5.4   Respiratory disease mortality

To date, only three prospective studies have analysed the association between circulating 25(OH)D and respiratory disease mortality; with inverse associations observed within them all. The largest (497 respiratory disease deaths) was from the UK based Whitehall II study, where a HR of 0.69 (95% CI: 0.56-0.85) was observed for a twofold higher measure of circulating 25(OH)D (61). Two recent smaller cohort studies also observed inverse associations between circulating 25(OH)D and respiratory disease mortality (59;62). Firstly, in a Danish cohort analysis in which 47 respiratory deaths were recorded (Q4 vs. Q1, HR 0.26, 95% CI: 0.09-0.75; *P*-trend 0.0042) (62). Most recently, a prospective German analysis reported reduced respiratory disease mortality when participants with the highest and lowest

levels of circulating 25(OH)D were compared (>50 vs. <30 nmol/L, HR 0.40, 95% CI: 0.18-0.89) (59). Generally, further larger prospective analyses are required to investigate the circulating 25(OH)D-respiratory disease mortality relationship.

1.6 Previous circulating 25-hydroxyvitamin D predictor scores

While directly measuring circulating 25(OH)D is viewed as the "gold standard" approach in prospective epidemiological studies to assess vitamin D-disease associations, the availability, collection and laboratory analysis of blood samples on a large-scale may be prohibitively expensive. Instead case-control studies nested within prospective cohorts are usually undertaken. However, nested case-control datasets are specific for a disease or disease subtype as the control participants, usually selected using incidence density sampling, are uniquely matched to cases by follow-up time, time of year of blood collection, age, and other criteria. This means that the relationships between vitamin D and other disease end-points cannot be carried out without creating a new disease-specific nested case-control dataset.

A further weakness of studies which directly measure circulating 25(OH)D is that blood samples have usually only been collected on one occasion (usually at baseline of the study). This means that within-person variation in circulating levels of 25(OH)D (half-life of ~3 weeks) is usually not taken into account. Vitamin D status at a given point in time is reflective of recent sun exposures and behaviours (e.g. beach holiday, season, sunbathing habits), as well as dietary/supplementary intakes. Correlations between circulating 25(OH)D levels within individuals has been shown to be fairly stable over 2-3 years (correlation coefficients of ~0.70) (52;162). However, studies which have investigated the reproducibility of circulating 25(OH)D levels over longer time periods (5 years plus) have observed attenuations in correlation coefficients to ~0.50 (163;164), suggesting that a single measure of circulating 25(OH)D may not be an optimal reflection of long-term vitamin D status.

An alternative approach to measuring actual circulating 25(OH)D, which has previously been used to assess disease risks in the U.S. based HPFS and Framingham Offspring cohorts, is to derive predicted circulating 25(OH)D scores (52;165). In both studies, actual circulating 25(OH)D measurements - available for a subset of cohort participants - were modelled in multiple regression models with predictors/correlates of vitamin D status, such as: location of residence, vitamin D intake, physical activity, and body size. Then the validated predictor scores were applied to the full cohorts, creating a predicted circulating 25(OH)D value for each participant. This variable was then used to assess risks of either cancer incidence and mortality (52) or type-2 diabetes (165). In validation analyses in the NHS and HPFS cohorts, similar associations for colorectal, pancreatic and prostate cancers, type-2 diabetes and hypertension were observed when the predicted 25(OH)D and actual circulating 25(OH)D measurements were used to assess vitamin D status (164). This approach is cost-effective as actual circulating 25(OH)D measurements to derive the predictor score are only required from a subset of cohort participants. Furthermore, once the scores have been created they can be applied to all cohort participants (minus those whose samples were used in predictor score derivation), meaning that multiple disease end-points can be assessed. Another advantage is that predictor scores may provide more stable long-term indicators of vitamin D status, than actual measurements of circulating 25(OH)D, as the constituent model variables used to derive the scores are relatively stable over time (e.g. BMI or region of residence) (9;52). Predictor scores of 25(OH)D status have only been previously used twice in U.S. cohorts to assess chronic disease risk. Whether this analytical approach is appropriate for European populations is unknown.

1.7   Other prediction scores derived in prospective cohort studies

While predictor scores of 25(OH)D have been previously used to estimate vitamin D exposures, most previous prediction scores derived from prospective cohort data have been used to directly predict dichotomous chronic disease outcomes. Most notably, the Framingham cardiovascular disease risk score (derived from the Framingham Heart Study) was developed to identify individuals at high-risk of developing the disease (166). Exposures included when calculating an individual's Framingham score are: age, blood cholesterol levels, smoking status and duration, blood pressure, and the presence of diabetes (166). In

the UK, the Framingham score has been adopted into clinical practice, as a patient's risk of a cardiovascular event (within 10 years) can be calculated in all electronic patient record systems (167).

Multiple prediction scores have also been derived from prospective cohort data for type 2 diabetes risk (167). The components of these scores usually were: age, anthropometric measurements, family history of diabetes, smoking status, hypertension, blood triglyceride and cholesterol levels, and in some scores, diet, alcohol consumption and physical activity (167). To date, none of the type 2 diabetes prediction scores have been used in a clinical setting to identify high and low-risk individuals (167).

1.8   Summary

Despite the large amount of recent research that has investigated the associations between vitamin D and cancer incidence and mortality, cause-specific mortality, and all-cause mortality, many uncertainties remain. The best evidence for a reduced incidence of disease is for colorectal cancer, while for other cancer types, heterogeneous results have emerged. For prostate cancer, previous research is non-supportive of an inverse association, with some recent studies even reporting an elevated risk with higher circulating 25(OH)D levels. While for breast cancer, the prospective data does not support a role for vitamin D in lowering disease risk. Other rarer cancers have been studied less frequently or not at all (e.g. thyroid cancer and liver cancer).

For other chronic disease end-points, such as circulatory disease mortality, prospective studies which measured circulating 25(OH)D levels have usually reported inverse associations; however, these studies were relatively small in terms of the number of recorded deaths. For respiratory disease mortality and digestive disease mortality few previous studies have been carried out.

Many of these previous studies contain small numbers of incidence cases/deaths, and have adjusted for other risk factors which may confound the vitamin D-cancer relationships to varying extents. Measurements of actual and predicted 25(OH)D levels may be especially vulnerable to residual confounding from other chronic disease risk factors as higher 25(OH)D levels are usually correlated with lower adiposity/BMI, higher physical activity levels, younger age, and not smoking (as detailed in Section 1.2). Due to these concerns, analyses in cohorts with a large number of recorded incident cases and deaths, and in which information on possible confounding variables have been extensively collected, are required.

More data are also required to investigate the unexpected elevated risks previously observed for prostate cancer incidence, pancreatic cancer incidence and total cancer mortality associated with higher levels of circulating 25(OH)D. Similarly, for the vitamin D-all-cause mortality relationship, the J- and U-shaped associations observed in previous studies require further investigation. Understanding the relationships between circulating 25(OH)D and chronic disease and all-cause mortality end-points is paramount in setting dosage regimens for future intervention studies and ultimately shaping public health policy. Further prospective research is required to aid understanding of these relationships.

Prospective studies directly measuring actual circulating 25(OH)D are viewed as the "gold standard" approach to assess the vitamin D associations. However, these studies are expensive to carry out (as circulating 25(OH)D has to be measured in all participants), disease specific (when used in nested case-control studies), and the single measurement of circulating 25(OH)D may not reflect long-term exposures. An alternative approach, not yet used in European populations, is to create predictor scores for circulating 25(OH)D. This cost effective approach means associations between predicted 25(OH)D and multiple outcomes (including rarer diseases) can be investigated.

This EPIC research used circulating 25(OH)D measures taken from a subset of the cohort to derive and validate predictor scores of 25(OH)D status. These predictor scores were then applied to the full EPIC cohort (minus those individuals whose circulating samples were used to derive the score) and used to assess risks of cancer incidence and mortality, circulatory disease mortality, respiratory disease mortality, digestive disease mortality, and all-cause

mortality. Prior to predictor score derivation, the relationships between dietary vitamin D and all-cause and cancer caused mortality within the full EPIC cohort were also assessed.

1.9   Aims

The aims of these analyses were: (1) to assess the relationships between dietary vitamin D and all-cause and cancer caused mortality in EPIC; (2) to derive and validate predictor 25(OH)D scores using correlates/determinants of vitamin D status; (3) to apply the validated predictor 25(OH)D scores to the full EPIC cohort to assess the incidence risks of overall cancer and individual cancers; (4) to apply the validated predictor 25(OH)D scores to the full EPIC cohort to assess the risks of cancer mortality, circulatory disease mortality, respiratory disease mortality, digestive disease mortality, and all-cause mortality.

2. METHODS

2.1 Dietary vitamin D intake and all-cause and cause-specific mortality

2.1.1 Study population

EPIC is an on-going multicentre prospective cohort study designed to investigate the association between diet, lifestyle, genetic and environmental factors and various types of cancer. A detailed description of the methods employed has previously been described (168;169). In summary, 521,448 participants (~70% women) mostly aged 35 years or above were recruited between 1992 and 2000. Participants were recruited from 23 study centres in ten European countries: Denmark (Aarhus and Copenhagen); France; Germany (Heidelberg and Potsdam); Greece; Italy (Florence, Naples, Ragusa, Turin, and Varese); the Netherlands (Bilthoven and Utrecht); Norway (Tromso); Spain (Asturias, Granada, Murcia, Navarra, and San Sebastian); Sweden (Malmö and Umea); and the United Kingdom (UK; Cambridge and Oxford).

Participants were recruited from the general population of their respective countries, with the following exceptions: the French cohort were teacher health insurance programme members; the Italian and Spanish cohorts included members of blood donor associations and the general population; the Utrecht (the Netherlands) and Florence (Italy) cohorts contained participants from mammographic screening programs; the Oxford (UK) cohort included a large proportion of vegetarians, vegans, and low meat eaters; finally, only women participated in the cohorts of France, Norway, Naples (Italy) and Utrecht (the Netherlands). Written informed consent was provided by all study participants. Ethical approval for the EPIC study was provided from the review boards of the International Agency for Research on Cancer (IARC) and local participating centres.

Exclusions prior to the onset of the analyses, included: participants with missing dietary vitamin D intake (n=6,837); participants in the highest and lowest 1% of the distribution for the ratio between energy intake to estimated energy requirement (n=10,242); participants

who had cancer at baseline (n=23,412); participants who self-reported a history of heart disease (n=11,163), stroke (n=3,258), or diabetes (n=11,438) at baseline; and finally, participants with missing follow-up information (n=2,381). This analysis therefore included 452,717 participants (130,564 men and 322,153 women).

2.1.2 Diet, lifestyle, and anthropometric information collection

Dietary information over the previous 12 months was obtained at study baseline using validated country/centre specific dietary questionnaires. In Malmö (Sweden), a dietary questionnaire was combined with a 7-day food registration and interview. In Greece, two Italian centres, and Spain, interviewers administered the dietary questionnaires. In all other centres/countries, the questionnaires were self-administered. In Spain, France, and Ragusa (Italy) questions were structured by meals, while in other countries the structure was by food groups. Intakes of dietary vitamin D were obtained from the EPIC Nutrient Data Base (ENDB); in which the nutritional composition of foods across the different countries has been standardised (170).

Lifestyle questionnaires were used to obtain information on education (used as a proxy for socioeconomic status), smoking habits (status, intensity and duration), alcohol consumption, and physical activity levels. Height and weight were measured at the baseline examination in all centres apart from part of Oxford and all of the Norway and France cohorts, where measurements were self-reported via the lifestyle questionnaire (168;169). BMI (kg/m^2) was calculated from these height and weight measures. Medication and information on reproductive history was obtained via questionnaires. Menopausal status at enrolment was calculated from an algorithm using information on menstrual history, menopause type, oral contraceptive and menopausal hormone use.

2.1.3 Assessment of mortality

Vital statuses, causes, and dates of death were obtained from record linkages with cancer

registries, boards of health, and death indexes (Denmark, Italy, the Netherlands, Norway, Spain, Sweden, and the United Kingdom). For Germany, Greece, and France participants were actively followed-up through a combination of methods: by mail or telephone directly; or through municipal registries, regional health departments, physicians, and hospitals. Data on causes of deaths were coded in accordance with the International Classification of Diseases, 10th Revision (ICD-10). Because of time differences across participating centres in reporting the causes of deaths, follow-up dates were truncated to when 80% of causes were known. Truncated follow-up dates were: 30 June 2005 for Cambridge (UK); 31 December 2006 for Denmark, France, Naples (Italy), Turin (Italy), Varese (Italy), Granada (Spain), Murcia (Spain), and Malmö (Sweden); 31 December 2007 for Florence (Italy), Norway, San Sebastian (Spain), and Umea (Sweden); 31 December 2008 for the Netherlands, Ragusa (Italy), Asturias (Spain), and Navarra (Spain); 30 June 2009 for Oxford (UK); and the actual date of last contact for Germany and Greece. Where the cause of death was coded with the qualifier "underlying" this was taken as the originating cause of death. Where one cause of death was given, this was used as the originating cause of death. If two or more recorded causes of death were given, with one being "antecedent" and one of the others being "immediate", the former was used as the cause of death. Finally, if deaths were classified with two or more causes including: "other significant conditions"; "not distinguished"; or "immediate", the latter cause was used. The underlying causes of death were used to estimate the risks of the following causes of death: cancer (ICD-10: C00-D48), circulatory diseases (I00-I99), respiratory diseases (J30-J98), and digestive diseases (K00-K93). Cancer deaths were further divided into digestive and non-digestive system cancers, based on the *a priori* hypothesis that the former would be more sensitive to vitamin D status. Digestive system cancers were oesophagus (C15); stomach (C16); colorectal cancer (C18-C20); anus and anal canal (C21); liver and intrahepatic bile acids (C22); gallbladder (C23); other and unspecified parts of biliary tract (C24); pancreas (C25); and other ill-defined digestive organs (C26). Non-digestive system cancers were all other cancers. Deaths caused from colorectal cancer (C18-C20), pancreatic cancer (C25), lung cancer (C34), prostate cancer (C61), and breast cancer (C50) were also investigated. Mortality risk caused by external causes (S00-Y98), such as accidents, was examined as negative controls.

2.1.4 Statistical analysis

Hazard ratios and 95% confidence intervals (CIs) were estimated using Cox proportional

hazards models; with age as the primary time variable in all models. Time at entry was age at recruitment. Exit time was age at death or the last date at which follow-up was considered complete in each centre. Models were stratified by study centre to control for differing follow-up procedures, questionnaire design, and other differences across centres. Models were also stratified by sex and age at recruitment in 1-year categories. Possible non-proportionality was assessed using an analysis of Schoenfeld residuals (171); with no evidence of non-proportionality detected.

Dietary intakes of vitamin D were modelled using sex-specific quintiles and as continuous variables (HR expressed per increment of 2.5 µg/day). Trend tests across quintiles were calculated by assigning the median value of each intake quintile and modelling as continuous terms into Cox regression models. Analyses were conducted for both sexes combined and separately. All models were adjusted for total energy intake, using the standard model, to obtain isocaloric risk estimates and partly control for measurement error of vitamin D intake estimates. All models were additionally adjusted for: BMI (<22, 22-24.9, 25-29.9, 30-34.9, or 35+ kg/m^2); physical activity index (inactive, moderately inactive, moderately active, active, or missing); smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 25+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; or unknown); smoking duration (<10, 10-<20, 20-<30, 30-<40, 40+ years, or smoking duration unknown); education level (none/primary school completed, technical/professional school, secondary school, longer education - including university, or unknown); menopausal status (premenopausal, postmenopausal, perimenopausal/unknown menopausal status, or surgical postmenopausal); ever use of oral contraceptive (yes, no, or unknown); ever use of menopausal hormone therapy (yes, no, or unknown); and intakes of alcohol (non-consumers, <5, 5-14.9, 15-29.9, or 30+ g/day), red and processed meats (continuous, g/day), fruits and vegetables (continuous, g/day), dietary calcium (continuous, mg/day), and polyunsaturated fatty acids (continuous, g/day). The colorectal cancer mortality analyses were additionally adjusted for cereal fibre intake (continuous, g/day); whilst the breast cancer mortality analysis was additionally adjusted for age at menarche (<12, 12-<15, 15+ years old, or unknown) and age at first pregnancy (<21, 21-<30, 30+ years old, no children, or not specified).

To determine whether the all-cause mortality association differed according to anthropometric and lifestyle characteristics, interaction terms (multiplicative scale) were

included in the models. The statistical significance of the cross-product terms were evaluated using the likelihood ratio test. Interaction terms inputted into the statistical model were intakes of dietary vitamin D (continuous) with sex, smoking status (never, former, or current); BMI (underweight-normal weight = <25; overweight = 25-29.9; obese = 30+ kg/m^2); and physical activity index (inactive, moderately inactive, moderately active, active). The heterogeneity across countries was explored by taking a meta-analytic approach (172). How the associations differed according to length of follow-up time was also explored (<5, 5-10, or ≥10 years). Follow-up time of <5 years included the person-time and incident events within this time period only. Follow-up time for 5-10 and ≥10 years included only the person-time and incident events within these respective time periods. To evaluate possible reverse causality, cases diagnosed within the first 5 years of follow-up were excluded from the analyses.

Statistical tests used in the analysis were all two-sided and a *P*-value of <0.05 was considered statistically significant. All analyses were performed using Stata version 11.0.

2.2 Derivation and validation of European Prospective Investigation into Cancer and Nutrition 25-hydroxyvitamin D predicted score

2.2.1 Study population and data collection

The study involved secondary analysis of subjects circulating 25(OH)D measures from previous EPIC colorectal cancer (n=2,388) (67), prostate cancer (n=1,077) (97), lymphoma (n=2,248) (136) and breast cancer (n=1,395, all controls) (117) nested case-control studies. Cases for the nested case-control studies were sourced from the full EPIC cohort as described in Section 2.1.1. In the colorectal cancer study, cases were not selected from Norway or the Malmö (Sweden) centre (67). In the prostate cancer study, the women only cohorts were excluded (France, Norway, Utrecht (the Netherlands), and Naples (Italy)); cases were also not selected from the Denmark and Malmö (Sweden) cohorts (97). In the lymphoma nested case-control analysis, the France cohort was excluded (136). Only control samples were available for predictor score derivation from the breast cancer nested case-control study (117). In all studies, controls were match 1:1 with cases by age, sex, centre, timing of blood collection. Cases and controls in the colorectal and prostate cancer studies were additionally matched by fasting status of blood collection; whilst additional matching criteria in the breast cancer and lymphoma was length of follow-up and blood donor status (lymphoma study only).

Participants with incomplete dietary intake information were excluded (n=24). Due to the incompatibly of the physical activity index information with the remainder of the cohort, participants from the Umea centre were excluded (n=310). Participants with missing data for physical activity index (n=134), smoking status and intensity (n=37), and waist circumference (n=485) were also excluded from the predictor score datasets. This meant that 6,118 participants (2,966 men and 3,152 women) with circulating 25(OH)D measurements were used for the derivation and validation of the predictor score.

Information on dietary intakes, physical activity, smoking status, education level and alcohol intake was obtained from participants as described in Section 2.1.2. Dietary intakes were adjusted for energy using the residual method (173). Additionally, waist circumference was

measured at recruitment from participants at ether the narrowest circumference of the torso or at the mid-point between ribs ad iliac chest.

2.2.2   Blood collection and laboratory measures

Blood samples were collected from 385,747 of the 521,448 EPIC participants at recruitment, prior to disease diagnosis. Samples were stored in liquid nitrogen tanks at IARC (Lyon, France) at -196°C, except for samples from Denmark (-150°C, nitrogen vapour). Circulating 25(OH)D levels for participants in the prostate cancer study were determined by enzyme immunoassay (OCTEIA 25-Hydroxy Vitamin D kit; Immunodiagnostic Systems, Limited, Boldon, Tyne and Wear, UK) in the MRC Research Laboratories in Cambridge (97). Laboratory personnel were blinded as to case or control status of participants. The same enzyme immunoassay kit was used for colorectal cancer participant blood samples (67). These analyses were conducted at the laboratory for Health Protection Research, National Institute for Public Health and the Environment, the Netherlands. Circulating samples in the breast cancer and lymphoma studies were analysed with the fully automated IDS-iSYS 25(OH)D (Immunodiagnostic systems Ltd, Boldon, UK). For quantification of 25(OH)D in EDTA-samples a OCTEIA 25-hydroxyvitamin D enzyme immunoassay (IDS, Immunodiagnostic systems Ltd, Boldon, UK) was used (117;136).

2.2.3   Derivation and validation of predicted 25-hydroxyvitamin D scores

Derivation of predicted 25-hydroxyvitamin D scores

For each sex, random two-third subsets of participants with blood measurements (n=1,982 for men and n=2,107 for women) were used to investigate the relationship between circulating 25(OH)D and dietary, lifestyle, geographical, age at blood collection, and timing of blood collection variables, forming the basis of the predictor 25(OH)D scores. Blood samples from the remaining third of men and women participants (n=984 for men and n=1,045 for women) were used to validate the predictor scores.

Multiple linear regression modelling was used with circulating 25(OH)D as the dependent variable and potential correlates/determinants as independent variables. To meet the model assumptions of a linear relationship between independent and dependent variables, circulating measurements of 25(OH)D were naturally log-transformed. When multiplied by 100 the $\log_e$ β coefficients from the linear regression analysis can be interpreted as the mean percentage difference compared with the reference category (categorical variables), or as the mean percentage difference in circulating 25(OH)D for a one unit increase in the predictor variable (continuous) (174).

Independent variables/correlates/determinants considered for inclusion in the predictor scores, in addition to sex, were chosen based on a prior knowledge of their relationships with circulating 25(OH)D. These were split into the following categories: age at blood collection and timing of blood collection, dietary variables, anthropometric variables, lifestyle variables, reproductive variables, and location of residence.

Age at blood collection and timing of blood collection variables
Variables considered for inclusion in the predictor scores were: age at blood collection (years; continuous); month of blood collection (January, February, March, April, May, June, July, August, September, October, November, or December); season of blood collection (spring, summer, autumn, or winter); and year of blood collection (1992, 1993, 1994, 1995, 1996, 1997, 1998, or 1999).

Dietary variables
Variables considered for inclusion were: dietary intakes of vitamin D (µg/day; continuous), calcium (mg/day; continuous), and retinol (µg/day; continuous).

Anthropometric variables
Variables considered as indicators of adiposity were: waist circumference (cm; continuous), BMI (kg/m^2; continuous); height (cm; continuous); weight (kg; continuous); and waist-to-hip ratio (continuous).

Lifestyle variables

Variables considered for inclusion were: physical activity index level (inactive, moderately inactive, moderately active, or active); recreational physical activity (METs; continuous); smoking status and intensity (never; current, 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 25+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; or unknown); education level (no education, primary school completed, technical/professional school, secondary school, longer education including university, or not specified); and alcohol consumption (g/day; continuous).

Reproductive variables

Variables considered for inclusion in the women's predictor score were: menopausal status (premenopausal, postmenopausal, perimenopausal/unknown menopausal status, or surgical postmenopausal); ever use of menopausal hormone therapy (yes, no, or unknown); ever use of contraceptive pill (yes, no, or unknown); and whether women participants had live born children (no or yes).

Location of residence

Country of residence was included in the predictor scores (Denmark, Germany, Greece, Italy, the Netherlands, Spain, or the UK).

Univariate linear regression models were fit for each of these independent variables. All variables with P-values <0.05 were then included in multivariable linear regression models; these models were adjusted for age at blood collection, and source study and batch of circulating sample analysis. One at a time, non-significant variables with the largest P-values (≥0.05) were removed from the multivariable models, until all remaining predictors were statistically significant. Also, taken into consideration at this stage was the amount of additional variance in $\log_e$ 25(OH)D that was explained by the inclusion of predictor variables in the models. A significant sex interaction (P-value <0.001) was observed and as a consequence separate predictor scores were derived for men and women

The predictor scores assumed that: $\log_e 25(OH)D = \beta_0 + \beta_i X_i + \beta_i X_i \ldots$, where β_0 represents the intercept and β_i represents coefficient values associated with the value of the independent variable 25(OH)D predictor, X_i.

The final correlates/determinants included within the separate men and women's predictor scores were: age at blood collection, dietary vitamin D intake, waist circumference, physical activity index, month of blood collection, country of residence, smoking status and intensity, ever use of menopausal hormone therapy, and source study and batch of serum samples. The rationale for the inclusion of these correlates/determinants (and the exclusion of the other variables considered) is outlined in the results section. The final circulating 25(OH)D predictor scores for men and women explained 34% and 28% of the overall variance in circulating 25(OH)D respectively (Table 11).

Age at blood collection and smoking status and intensity were not used when the predictor scores were applied to the validation datasets (i.e. used only as covariates in predictor score derivation); as both are major risk factors for chronic disease incidence and mortality, and excluding them at this stage meant that they could be used as confounders when assessing disease risk (164). This minimised risk of statistical over-adjustment of these variables on the analyses. The timing of blood collection reflects recent sun or dietary exposures, and does not determine long-term average between person variation in circulating 25(OH)D, which is of interest when assessing disease risk; because of this month of blood collection was also excluded when the predictor score was applied. The predictor scores were also adjusted for source study and batch of circulating sample analysis as the serum samples were sourced from four different studies and analysed at different times and in different laboratories.

Validation of predicted 25-hydroxyvitamin D scores

The predictor scores were validated primarily using the blood samples from the remaining third of men and women participants (n=984 for men and n=1,045 for women). The predictor scores were applied in the validation datasets with actual 25(OH)D measurements, creating a predicted circulating 25(OH)D value for each participant. The various validation stages are outlined below:

Correlations between predicted 25-hydroxyvitamin D and actual circulating 25-hydroxyvitamin D

Firstly, Pearson and Spearman correlation coefficients (adjusted by source study and batch of circulating samples) were used to assess agreement between actual and predicted circulating 25(OH)D values.

Actual 25-hydroxyvitaminD measurements by quantile of predicted 25-hydroxyvitamin D

Next, the actual circulating 25(OH)D measurements according to quintile and decile of predicted 25(OH)D scores for men and women were examined.

Cross-classifications of participants by predicted and actual circulating 25-hydroxyvitamin D categories

Participants were also cross-classified by quintiles, tertiles and three or five biologically relevant pre-defined categories (three: <50 nmol/L deficiency, 50-<75 nmol/L insufficient, and 75+ nmol/L sufficient; five: <25 nmol/L severe deficiency, 25-<50 nmol/L deficiency, 50-<75 nmol/L insufficient, 75-<100 nmol/L sufficient, and 100+ nmol/L optimal) of both predicted and actual circulating 25(OH)D.

Assessment of colorectal cancer risk in the nested case-control dataset

The next validation stage was to assess colorectal cancer risk using the predictor scores in the colorectal nested case-control dataset; to find out if a similar inverse association was observed to the published analysis which used actual circulating 25(OH)D measurements (67). Firstly, to ensure independence of circulating 25(OH)D measures, participants whose serum samples were sourced from the colorectal cancer nested case-control study were excluded from the datasets used to derive the predictor scores (this left n=1,234 samples for men and n=1,372 samples for women). The predictor scores were then re-derived and applied to the colorectal cancer nested case-control dataset. In this dataset the controls were matched to cases on a 1:1 basis, by: age at recruitment; sex; study centre; time of day of blood collection; and women were further matched for menopausal status, phase of menstrual cycle, and usage of menopausal hormone therapy. Predicted circulating 25(OH)D was split into quintiles based on the distribution in the control participants within the dataset. Linear trends across these quintiles were assessed by assigning a score variable with

values from 1 to 5, dependent on quintile categorisation. Conditional logistic regression models were then used - stratified by case-control set - and adjusted for the same confounders as the published nested case-control analysis: BMI, smoking status and intensity, alcohol consumption, education, and intakes of total energy, fruits and vegetables, and meats or meat products. Multivariable models also included adjustment for physical activity index, although this inclusion may have caused statistical over-adjustment as physical activity index was also a key determinant of 25(OH)D included within the predictor scores. Due to this concern, separate risk estimates for multivariable models with and without physical activity index adjustment were assessed. Due to the close correlations between waist circumference (the marker of adiposity included within the predictor scores) and BMI, separate risk estimates for the multivariable model minus BMI adjustment were also assessed.

Assessment of colorectal cancer risk in the full European Prospective Investigation into Cancer and Nutrition cohort

The associations between predicted 25(OH)D and colorectal cancer were also assessed in the full EPIC cohort. Here the full datasets used to derive the predictor scores (containing n=1,982 men and n=2,107 women) were applied to the full EPIC cohort (minus participants with actual circulating 25(OH)D measurements who were used in the derivation and validation of predictor scores). Hazard ratios and 95% CIs were estimated using Cox proportional hazards models; with age as the primary time variable in all models. Time at entry was age at recruitment. Exit time was age at whichever of the following came first: cancer diagnosis, death, or the last date at which follow-up was considered complete in each centre. Predicted circulating 25(OH)D values were modelled using sex specific quintiles defined across cohort participants or as continuous variables (HR expressed per increment of 5.9 nmol/L, which is equivalent to 100 IU/L). Trend tests across quintiles of exposure were calculated by assigning the median value of each intake category and modelling as continuous terms into Cox regression models. These validation analyses used the same confounding variables as were controlled for in the published nested case-control analysis (67). The Cox regression models were stratified by age at recruitment, sex, and centre, and adjusted for BMI, total energy intake, education, smoking status and intensity, and intakes of alcohol, fruits and vegetables and meat and meat products. Once more, the multivariable model was also assessed with and without additional adjustment for physical activity index and BMI.

Assessment of prostate cancer risk in the full European Prospective Investigation into Cancer and Nutrition cohort

The predictor scores were also applied to the full EPIC cohort to assess prostate cancer risk. A non-significant suggestive positive association was observed in the published nested case-control study (97). The full dataset used to build the predictor score (containing n=1,982 men) was applied to the full EPIC cohort (minus participants with actual circulating 25(OH)D measurements who were used in the derivation and validation of the predictor score). The associations were assessed using Cox regression models, stratified by age at recruitment and centre, and adjusted for BMI, education, smoking status and intensity, marital status, and alcohol intake. Once more, the multivariable model was also assessed with and without additional adjustment for physical activity index and BMI. Predicted circulating 25(OH)D values were modelled using sex specific quintiles defined across cohort participants or as continuous variables (HR expressed per increment of 25 nmol/L). Trend tests across quintiles of exposure were calculated by assigning the median value of each intake category and modelling as a continuous term into Cox regression models.

2.3 Application and validation of the Health Professionals Follow-Up Study derived 25-hydroxyvitamin D score in the European Prospective Investigation into Cancer and Nutrition

2.3.1 Application of the Health Professionals Follow-Up Study derived 25-hydroxyvitamin D score in the European Prospective Investigation into Cancer and Nutrition

To test whether prediction models derived in different cohorts can be applied to other study populations, we applied the HPFS predictor score of Giovannucci *et al.,* (52) to men in EPIC with circulating 25(OH)D measurements available. The HPFS predictor score included the following correlates/determinants of circulating 25(OH)D: race; residence location; quintile of leisure-time physical activity; BMI; dietary vitamin D intake; supplementary vitamin D intake; and season of blood collection.

Residence in the HPFS predictor score was split into three groups: South, Midwest/West, and Northeast/Mid-Atlantic. The latitudes of these locations span from approximately 37°N to 47°N; whilst the latitude range of EPIC constituent countries is approximately 36°N to 63°N. Participants in EPIC countries located outside the HPFS latitude ranges were excluded from the analysis, these were: Denmark (n=472), Germany (n=683), the Netherlands (n=79), and the UK (n=684). This meant that 823 circulating 25(OH)D measurements were available for application of the HPFS score within EPIC. The remaining centres were split into southern and northern European categories, with equivalent HPFS coefficients for Northeast/Mid-Atlantic and the South used for these categories. The centres categorised as the southern Europe reference category were: Ragusa (36°N; Italy), Granada (37°N; Spain), Greece (37°N, Athens), Murcia (37°N; Spain), Navarra (42°N; Spain), Asturias (43°N; Spain), and San Sebastian (43°N; Spain). Centres categorised as northern European were Florence (43°N; Italy), Turin (45°N; Italy), and Varese (45°N; Italy). For quintiles of leisure-time physical activity, the EPIC recreational physical activity variable split into quintiles was used. Dietary vitamin D was categorised into five intake groups: ≥400; 200-399; 100-199; and <100 IU/day. Three HPFS determinants were excluded when the predictor score was applied within EPIC. Firstly, supplementary vitamin D was not included as intake information was unavailable for EPIC participants. Secondly, race was excluded as participants within

EPIC are believed to be all Caucasian. Finally, season was excluded as this timing of blood collection reflects recent exposures and not long-term between person variation in 25(OH)D. This meant that final HPFS predictor score for application within EPIC included four determinants of circulating 25(OH)D: residence, quintile of leisure-time/recreational physical activity, BMI, and dietary vitamin D intake (Table 3).

Table 3. Correlates/determinants of circulating 25-hydroxyvitamin D included in the Health Professionals Follow-up study (52) and the adapted coefficient values used when applied within the European Prospective Investigation into Cancer and Nutrition cohort

Determinants	Change in 25(OH)D (nmol/L) as per HPFS	Change in 25(OH)D (nmol/L) adapted for EPIC	N
Intercept	90.8	90.8	-
Residence β			
South	REF	REF	283
Midwest/West	−2.4	-	-
Northeast/Mid-Atlantic	−6.4	-6.4	540
Quintile of leisure-time physical activity Ω			
5	REF	REF	133
4	−4.5	−4.5	157
3	−7.7	−7.7	125
2	−9.0	−9.0	176
1	−13.5	−13.5	232
Body mass index (kg/m^2)			
<22	REF	REF	28
22–24.9	−1.0	−1.0	148
25–29.9	−4.5	−4.5	448
30–34.9	−6.5	−6.5	168
≥35	−8.6	−8.6	31
Dietary vitamin D (IU/day) ‡			
≥400	REF	REF	22
300–399	−3.5	−3.5	28
200–299	−2.6	−2.6	104
100–199	−7.2	−7.2	316
<100	−10.4	−10.4	353
Supplementary vitamin D (IU/day) Φ			
≥400 IU/day	REF	REF	-
200–399 IU/day	−1.8	-	-
100–199 IU/day	2.4	-	-
<100 IU/day	−2.1	-	-
Race §			
White	REF	REF	823
African American	−12.8	-	-
Asian	−13.3	-	-
Season of blood collection ¥			
Autumn	REF	REF	-
Summer	−1.8	-	-
Spring	−12.1	-	-
Winter	−13.5	-	-

§ Participants within EPIC are solely Caucasian.

β Residence locations in U.S. based HPFS range from latitudes of ~37° - 47°; compared to EPIC pan-European latitude range ~36° - 63°. Greece, Italy, and Spain the only EPIC countries included.

Ω Recreational physical activity quintiles used within EPIC.

‡ Dietary vitamin D energy adjusted using the residual method, and converted into IU/day from µg/day (1 µg of vitamin D=40 IU).

Φ Dietary supplement information not available within EPIC.

¥ Season excluded when predictor score applied as not a factor in determining long-term between person variation.

2.3.2 Validation of the Health Professionals Follow-Up Study derived 25-hydroxyvitamin D score in the European Prospective Investigation into Cancer and Nutrition

The HPFS predictor score performance was validated within EPIC using a four stage process: (1) Pearson and Spearman correlation coefficients (adjusted by source study and batch of circulating samples) between actual and predicted circulating 25(OH)D values were calculated; (2) actual circulating 25(OH)D measurements according to decile of predicted 25(OH)D scores were examined; (3) participants were cross-classified by quintile of predicted and actual circulating 25(OH)D, and the agreement between these categorisations was assessed; and finally (4) assessment of colorectal cancer risk amongst men in the full EPIC cohort.

2.4 Application of predicted 25-hydroxyvitamin D scores to assess cancer incidence risk

2.4.1 Study population

The EPIC study population was sourced as described above (Section 2.1.1). Participants used in the derivation of the predictor score were excluded from the analyses (n=7,108). Also excluded were participants with missing information for any of the predictor score correlates/determinants: dietary vitamin D intake (n=6,821); physical activity index (n=44,507); waist circumference (n=78,027); and smoking status and intensity (n=2,926). Participants from Sweden were excluded as they were not included within the predictor score datasets (n=27,722). Other exclusions prior to the onset of the analyses, included: participants in the highest and lowest 1% of the distribution for the ratio between energy intake to estimated energy requirement (n=6,982); and participants with prevalent cancer at baseline (n=16,931). This analysis therefore included 330,424 participants (112,957 men and 217,467 women).

2.4.2 Diet, lifestyle, and anthropometric information collection

Dietary, lifestyle, and anthropometric information were collected from participants as detailed in Section 2.1.2.

2.4.3 Assessment of cancer incidence

Population cancer registries were used in Denmark, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom to identify incident cancer diagnoses. In France, Germany and Greece cancer cases during follow-up were identified by a combination of methods including: health insurance records, cancer and pathology registries, and by active follow-up directly through study participants or through next-of-kin. Complete follow-up censoring dates varied amongst centres, ranging between 2005 and 2010.

Cancer incidence data were coded using the 10th Revision of the International Classification of Diseases (ICD-10) and the second revision of the International Classification of Disease for Oncology (ICDO-2). Only the first primary neoplasm was included in the analyses. Cancer incidence end-points considered in this analysis were: overall cancer incidence (C00-C97); colorectal cancer (C18-C20); lung cancer (C34); kidney cancer (C64-C65); stomach and oesophageal cancers (C15-C16); bladder cancer (C67); pancreatic cancer (C25); liver cancer (C22-C24); brain cancer (C70- C72); skin cancer (C44); thyroid cancer (C73); prostate cancer (C61); breast cancer (C50); ovarian cancer (C48, C56-C57); endometrial cancer (C54); and non-Hodgkin's lymphoma (Originally classified according to ICD-O-2, but were reclassified according to the WHO classification of haematopoietic and lymphoid tissues, third edition (171). The conversion was performed using a program available on the United States National Cancer Institute Surveillance, Epidemiology and End Results (SEER) website (http://seer.cancer.gov/) and by the expertise of pathologists). Cancer cases were additionally split into digestive system (Colorectal, upper aero-digestive tract, stomach, oesophageal, pancreatic and liver cancers) and non-digestive system cancers (all other cancers). Similarly, cases were split into smoking related (Lung, bladder, upper aero-digestive tract, kidney, stomach, oesophageal, pancreatic, liver, and colorectal cancers) and non-smoking related cancers (all other cancers).

2.4.4 Statistical analysis

Hazard ratios and 95% confidence intervals (CIs) were estimated using Cox proportional hazards models; with age as the primary time variable in all models. Time at entry was age at recruitment. Exit time was age at cancer diagnosis, or the last date at which follow-up was considered complete in each centre. To control for differing follow-up procedures, questionnaire design, and other differences across centres, models were stratified by study centre. Models were also stratified by sex and age at recruitment in 1-year categories. Possible non-proportionality was assessed using an analysis of Schoenfeld residuals (171); with no evidence of non-proportionality detected.

Predicted circulating 25(OH)D exposures were modelled using sex-specific quintiles and as continuous variables (HR expressed per increment of 25 nmol/L). Trend tests across

quintiles were calculated by assigning the median value of each intake quintile and modelling as continuous terms into Cox regression models. Analyses were conducted for both sexes combined and separately. All models were adjusted for: BMI (<22, 22-24.9, 25-29.9, 30-34.9, or 35+ kg/m^2); smoking status and intensity (never; current, 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 25+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; or unknown); smoking duration (<10, 10-<20, 20-<30, 30-<40, 40+ years, or smoking duration unknown); education level (none/primary school completed, technical/professional school, secondary school, longer education - including university, or unknown); menopausal status (premenopausal, postmenopausal, perimenopausal/unknown menopausal status, or surgical postmenopausal); ever use of oral contraceptive (yes, no, or unknown); and intakes of total energy (continuous, kcal/day), alcohol (non-consumers, <5, 5-14.9, 15-29.9, or 30+ g/day), red and processed meats (continuous, g/day), fruits and vegetables (continuous, g/day), and dietary calcium (continuous, mg/day). The colorectal cancer analyses were additionally adjusted for cereal fibre intake (continuous, g/day); whilst the breast cancer, ovarian cancer, and endometrial cancer analyses were additionally adjusted for age at menarche (<12, 12-<15, 15+ years old, or unknown) and age at first pregnancy (<21, 21-<30, 30+ years old, no children, or not specified). Physical activity index was an important correlate/determinant in the 25(OH)D predictor scores, and also an important risk factor/confounder for cancer. Inclusion of physical activity index in both the predictor score and as a confounder in the mortality models may be deemed as statistical over-adjustment. Therefore, risk estimates for all models were presented with and without this additional adjustment for physical activity index (inactive, moderately inactive, moderately active, active, or missing). Similarly, due to the close correlations between waist circumference (the marker of adiposity included within the predictor scores) and BMI, inclusion of the latter within the cancer incidence models could be deemed as statistical over-adjustment. Therefore, separate risk estimates for the multivariable model minus BMI adjustment are also presented in the appendix.

To determine whether cancer incidence associations differed according to anthropometric, lifestyle, and dietary characteristics, interaction terms (multiplicative scale) were included in the models. The statistical significance of these cross-product terms was evaluated using the likelihood ratio test. Interaction terms inputted into the statistical model were predicted circulating 25(OH)D (continuous) with: sex; age at recruitment (<55, 55-<65 or ≥65 years old); smoking status (never, former, or current); BMI (underweight-normal weight = <25; overweight = 25-29.9; obese = 30+ kg/m^2); physical activity index (inactive, moderately

inactive, moderately active, or active); alcohol consumption (non-consumers, <15, 15-29.9, or ≥30 g/day); menopausal status (premenopausal, postmenopausal, perimenopausal, or surgical postmenopausal); ever use of the contraceptive pill (yes/no); and above and below median dietary intakes of red and processed meat (median=68.6 g/day), fruits and vegetables (median=412 g/day), fish (median=23.5 g/day), calcium (median=959 mg/day), fibre (median=22.4 g/day), and retinol (median=557 µg/day). How the associations differed according to length of follow-up time was also explored (<5, 5-10, or ≥10 years). Follow-up time of <5 years included the person-time and incident events within this time period only. Follow-up time for 5-10 and ≥10 years included only the person-time and incident events within these respective time periods. To evaluate possible reverse causality, cases diagnosed within the first 5 years of follow-up were excluded from the analyses. To be able to determine the relative importance of the predictor score correlates/determinants on the risk of overall cancer incidence, mutually adjusted analyses for dietary vitamin D, physical activity index, waist circumference, country, and ever use of menopausal hormone therapy were conducted. These models were also adjusted for the same covariates as the main cancer incidence models. The heterogeneity across countries was explored by taking a meta-analytic approach (172).

2.5 Application of predicted 25-hydroxyvitamin D scores to assess all-cause and cause specific mortality risk

2.5.1 Study population

The EPIC study population was sourced as described above (Section 2.1.1). Participants used in the derivation of the predictor score were excluded from the analyses (n=7,108). Also excluded were participants with missing information for any of the predictor score correlates/determinants: dietary vitamin D intake (n=6,821); physical activity index (n=44,507); waist circumference (n=78,027); and smoking status and intensity (n=2,926). Participants from Sweden were excluded as they were not included within the predictor score (n=27,722). Other exclusions prior to the onset of the analyses, included: participants in the highest and lowest 1% of the distribution for the ratio between energy intake to estimated energy requirement (n=6,086); participants who had cancer at baseline (n=12,994); participants who self-reported a history of heart disease (n=9,684), stroke (n=2,680), or diabetes (n=9,240) at baseline; and finally, participants with missing follow-up information (n=2,896). This analysis therefore included 310,757 participants (103,251 men and 207,506 women).

2.5.2 Diet, lifestyle, and anthropometric information collection

Dietary, lifestyle, and anthropometric information were collected from participants as detailed in Section 2.1.2.

2.5.3 Assessment of all-cause and cause-specific mortality deaths

Assessment of all-cause and cause specific mortality cases is detailed in Section 2.1.3. The underlying causes of death were used to estimate the risks of the following causes of death: cancer (ICD-10: C00-D48), circulatory diseases (I00-I99), respiratory diseases (J30-J98), and digestive diseases (K00-K93). Cancer deaths were further divided into digestive

and non-digestive system cancers, based on the *a priori* hypothesis that the former would be more sensitive to vitamin D status. Digestive system cancers were: oesophagus (C15); stomach (C16); colorectal (C18-C20); anus and anal canal (C21); liver and intrahepatic bile acids (C22); gallbladder (C23); other and unspecified parts of biliary tract (C24); pancreas (C25); and other ill-defined digestive organs (C26). Non-digestive system cancers were all other cancers. Cancer deaths were additionally split by smoking and non-smoking related cancers. Smoking related cancers were: oral cavity (C01-C06, C08); oropharynx (C09, C10, C12-C14); nasopharynx (C11); oesophagus (C15); stomach (C16); colorectal (C18-C20); liver (C22); pancreas (C25); nasal cavity and sinuses (C300, C31); larynx (C32); lung (C34); kidney (C64); bladder (C65, C67); and myeloid leukemia (C92). Non-smoking related cancers were all other cancers. Deaths caused from colorectal cancer (C18-C20), pancreatic cancer (C25), lung cancer (C34), skin cancer (C43-C44), prostate cancer (C61), and breast cancer (C50) were also investigated. Mortality risk caused by external causes (S00-Y98), such as accidents, was examined as negative controls.

2.5.4　Statistical analysis

Hazard ratios and 95% confidence intervals (CIs) were estimated using Cox proportional hazards models; with age as the primary time variable in all models. Time at entry was age at recruitment. Exit time was age at death or the last date at which follow-up was considered complete in each centre. To control for differing follow-up procedures, questionnaire design, and other differences across centres, models were stratified by study centre. Models were also stratified by sex and age at recruitment in 1-year categories. Possible non-proportionality was assessed using an analysis of Schoenfeld residuals (171); with no evidence of non-proportionality detected.

Predicted circulating 25(OH)D exposures were modelled using sex-specific quintiles and as continuous variables (HR expressed per increment of 25 nmol/L). Trend tests across quintiles were calculated by assigning the median value of each intake quintile and modelling as continuous terms into Cox regression models. Analyses were conducted for both sexes combined and separately. All models were adjusted for: BMI (<22, 22-24.9, 25-29.9, 30-34.9, or 35+ kg/m^2); smoking status and intensity (never; current, 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 25+ cigarettes per day; former, quit ≤10

years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; or unknown); smoking duration (<10, 10-<20, 20-<30, 30-<40, 40+ years, or smoking duration unknown); education level (none/primary school completed, technical/professional school, secondary school, longer education - including university, or unknown); menopausal status (premenopausal, postmenopausal, perimenopausal/unknown menopausal status, or surgical postmenopausal); ever use of oral contraceptive (yes, no, or unknown); and intakes of total energy (continuous, kcal/day), alcohol (non-consumers, <5, 5-14.9, 15-29.9, or 30+ g/day), red and processed meats (continuous, g/day), fruits and vegetables (continuous, g/day), and dietary calcium (continuous, mg/day). The colorectal cancer mortality analyses were additionally adjusted for cereal fibre intake (continuous, g/day); whilst the breast cancer analyses were additionally adjusted for age at menarche (<12, 12-<15, 15+ years old, or unknown) and age at first pregnancy (<21, 21-<30, 30+ years old, no children, or not specified).

Physical activity index was an important determinant of the circulating 25(OH)D predictor score and also an important confounder for chronic disease mortality. Inclusion of physical activity index in both the predictor score and as a confounder in the mortality models may be deemed as statistical over-adjustment. Therefore, risk estimates for all models were presented with and without additional adjustment for physical activity index (inactive, moderately inactive, moderately active, active, or missing). Similarly, due to the close correlations between waist circumference (the marker of adiposity included within the predictor scores) and BMI, inclusion of the latter within the mortality models could be deemed as statistical over-adjustment. Therefore, separate risk estimates for the multivariable model minus BMI adjustment are presented in the appendix.

To determine whether the mortality associations differed according to lifestyle, demographic, anthropometric, and dietary characteristics, interaction terms (multiplicative scale) were included in the models. The statistical significance of these cross-product terms was evaluated using the likelihood ratio test. Interaction terms inputted into the statistical model were predicted circulating 25(OH)D (continuous) with: sex; age at recruitment (<55, 55-<65 or ≥65 years old); smoking status (never, former, or current); BMI (underweight-normal weight = <25; overweight = 25-29.9; obese = 30+ kg/m^2); physical activity index (inactive, moderately inactive, moderately active, or active); alcohol consumption (non-consumers, <15, 15-29.9, or ≥30 g/day); menopausal status (premenopausal, postmenopausal, perimenopausal, or surgical postmenopausal); ever use of the contraceptive pill (yes/no);

and above and below median dietary intakes of red and processed meat (median=68.5 g/day), fruits and vegetables (median=411 g/day), fish (median=23.1 g/day), calcium (median=961 mg/day), fibre (median=22.4 g/day), and retinol (median=559.2 µg/day). Follow-up time of <5 years included the person-time and incident events within this time period only. Follow-up time for 5-10 and ≥10 years included only the person-time and incident events within these respective time periods. To evaluate possible reverse causality, cases diagnosed within the first 5 years of follow-up were excluded from the analyses. To be able to determine the relative importance of the predictor score correlates/determinants on the risk of death, mutually adjusted analyses for dietary vitamin D, physical activity index, waist circumference, country, and ever use of menopausal hormone therapy were conducted. These models were also adjusted for the same covariates as the main mortality models. The heterogeneity across countries was explored by taking a meta-analytic approach (172).

Statistical tests used in the analysis were all two-sided and a *P*-value of <0.05 was considered statistically significant. All analyses were performed using Stata version 11.0.

3. RESULTS

3.1 Dietary vitamin D and all-cause and cancer caused mortality

3.1.1 Characteristics of cohort participants

After a mean (SD) follow-up of 12.7 (2.4) years, 24,997 all-cause mortality death were recorded amongst the 452,717 participants. The total person-years and distribution of all-cause mortality deaths by country are shown in Table 4. The 130,564 men contributed 1,610,859 years of follow-up and 11,118 all-cause deaths. The 322,153 women contributed 4,140,966 years of follow-up and 13,879 all-cause deaths. Intakes of dietary vitamin D amongst men and women were relatively low in Italy and Greece; whilst high intakes were reported in Sweden, Spain, and Norway (women only). Mortality rates were calculated per 10,000 person-years and age adjusted to the European standard population (175). Amongst men and women, the age adjusted mortality rates were 98 and 62 cases per 10,000 person-years respectively.

Table 4. Cohort characteristics by sex and country of participants included in the dietary vitamin D-mortality analyses

Country	N of participants		Total person-years		N all-cause deaths		Mean (SD) total dietary vitamin D intake (µg/day) *	
	Men	Women	Men	Women	Men	Women	Men	Women
Denmark	23,864	27,331	276,093	323,395	2,518	1,808	5.6 (11.1)	3.8 (6.4)
France	-	65,605	-	978,205	-	2,833	-	2.7 (4.5)
Germany	19,225	26,288	216,129	298,323	1,188	655	4.1 (8.1)	3.2 (7.3)
Greece	9,355	13,646	87,710	135,843	760	518	3.4 (4.8)	2.6 (3.5)
Italy	13,424	29,517	168,539	358,788	580	814	2.5 (4.2)	1.8 (3.3)
Norway	-	34,517	-	378,323	-	739	-	4.2 (5.9)
Spain	13,968	23,362	188,597	319,849	932	611	6.1 (9.2)	3.9 (5.9)
Sweden	20,892	25,609	288,222	358,617	2,403	1,724	8.6 (6.1)	6.1 (4.7)
The Netherlands	9,272	25,767	121,748	337,610	443	1,395	5.6 (5.3)	3.8 (4.5)
United Kingdom	20,564	50,511	263,822	652,014	2,294	2,782	4.4 (4.4)	3.2 (4.0)
All EPIC	130,564	322,153	1,610,859	4,140,966	11,118	13,879	5.5 (7.7)	3.6 (5.2)

* Data are intake information collected from 24-hour dietary recalls (n=32,221 participants)

3.1.2 Dietary sources of vitamin D amongst cohort participants

Amongst men and women, the greatest proportion of dietary vitamin D was consumed from fish and shellfish (Figure 3). Other common dietary sources of vitamin D were added fats (such as margarines) and eggs and egg products.

Figure 3. Dietary sources of vitamin D in the European Prospective Investigation into Cancer and Nutrition
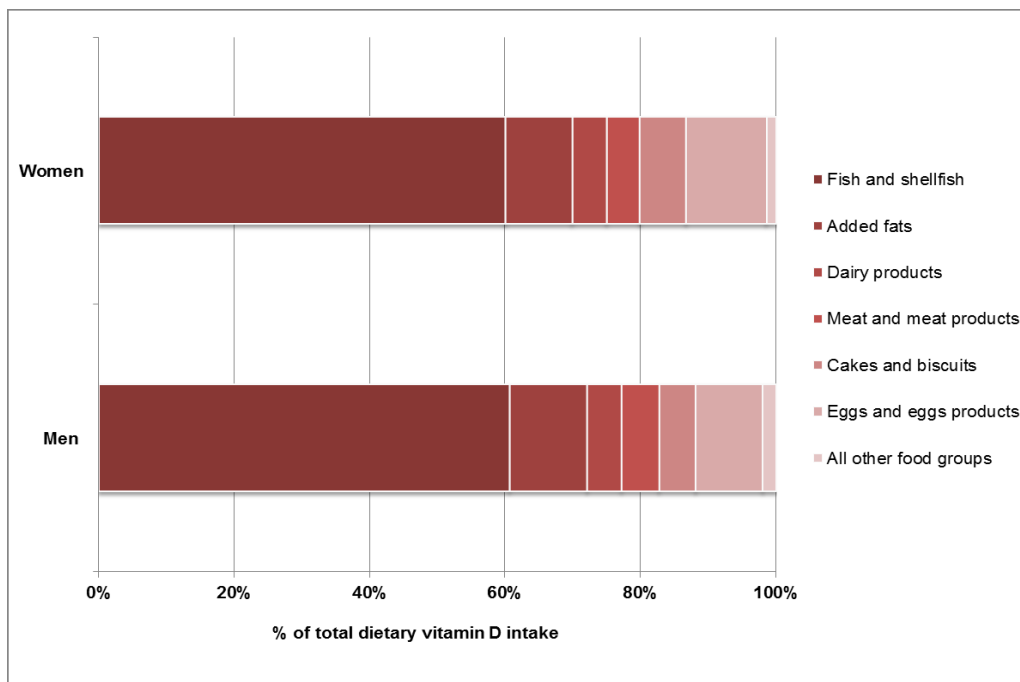
* Intake information collected from 24-hour dietary recalls (n=32,221 participants)

3.1.3 Baseline characteristics by dietary vitamin D intake quintiles

A higher proportion of current smokers were observed amongst men in the lowest and women in the highest vitamin D intake quintiles (Table 5). The lowest proportion of men classified as physically active was observed amongst those in the lowest vitamin D intake quintile. Compared to those in the lower intake quintiles, men and women with higher reported dietary vitamin D intakes also reported higher intakes of total energy, red and processed meats, calcium, and polyunsaturated fatty acids (Table 5).

Table 5. Baseline characteristics of study participants included in the mortality analyses by categories (quintiles) of total dietary vitamin D intake

Characteristic	Quintile of dietary vitamin D intake				
	Q1	Q2	Q3	Q4	Q5
Men					
Dietary vitamin D intake range (µg/day)	**<2.3**	**2.3-<3.4**	**3.4-<4.7**	**4.7-<6.5**	**≥6.5**
N	26,113	26,113	26,113	26,113	26,112
N all-cause deaths	1,996	1,835	2,154	2,353	2,780
Age at recruitment (years)	50.8 (10.9)	50.6 (9.9)	51.5 (9.7)	52.1 (9.5)	52.5 (10.0)
Body mass index (kg/m2)	26.6 (3.7)	26.4 (3.6)	26.4 (3.6)	26.3 (3.6)	26.3 (3.6)
Education ‡					
Longer education including University (%)	26.5	27.4	28.5	27.9	25.6
Smoking status and intensity ‡					
Current (%)	30.8	30.2	29.6	29.0	28.8
Physical activity ‡					
Active (%)	21.5	24.9	26.5	26.1	25.5
Total energy intake (kcal/day)	2094 (565)	2330 (626)	2412 (633)	2520 (633)	2783 (651)
Red and processed meat consumption (g/day)	67.8 (50.0)	92.5 (54.0)	104.2 (58.8)	108.6 (64.5)	114.1 (66.3)
Fruit & vegetable consumption (g/day)	511.8 (340.1)	410.9 (294.7)	356.8 (244.3)	344.9 (228.7)	359.3 (240.4)
Calcium intake (mg/day)	902.9 (356.7)	1013 (412.0)	1030 (422.7)	1078 (428.7)	1201 (457.4)
Polyunsaturated fat intake (g/day)	12.2 (6.1)	13.9 (6.3)	14.9 (6.3)	16.2 (6.7)	18.5 (7.2)
Alcohol intake (g/day)	19.6 (23.3)	21.5 (23.2)	21.4 (22.9)	20.8 (22.9)	19.3 (22.4)
Women					
Dietary vitamin D intake range (µg/day)	**<1.9**	**1.9-<2.7**	**2.7-<3.6**	**3.6-<5.1**	**≥5.1**
N	64,431	64,431	64,430	64,431	64,430
N all-cause deaths	2,699	2,516	2,761	2,866	3,037
Age at recruitment (years)	50.3 (10.9)	50.4 (9.6)	50.5 (9.4)	50.6 (9.4)	50.7 (9.2)
Body mass index (kg/m2)	25.0 (4.6)	24.6 (4.3)	24.7 (4.2)	25.0 (4.3)	25.0 (4.3)
Education ‡					
Longer education including University (%)	24.4	25.2	24.2	22.3	19.4
Smoking status and intensity ‡					
Current (%)	17.2	18.1	18.8	20.5	23.9
Physical activity ‡					
Active (%)	12.4	14.4	15.8	16.1	12.0
Total energy intake (kcal/day)	1657 (438)	1883 (492)	1981 (533)	2055 (558)	2100 (550)
Red and processed meat consumption (g/day)	47.7 (37.1)	64.7 (40.4)	70.3 (42.9)	72.8 (43.9)	69.9 (42.5)
Fruit & vegetable consumption (g/day)	507.7 (298.9)	467.2 (263.7)	452.1 (247.8)	444.4 (248.0)	432.1 (252.2)
Calcium intake (mg/day)	855.9 (346.9)	960.2 (376.2)	1003 (396.2)	1029 (415.1)	1015 (428.2)
Polyunsaturated fat intake (g/day)	10.3 (5.0)	11.9 (5.1)	12.9 (5.2)	13.8 (5.7)	14.9 (6.2)
Alcohol intake (g/day)	6.9 (11.1)	8.5 (12.0)	9.0 (12.2)	8.5 (12.1)	6.9 (10.2)
Ever use of contraceptive pill ‡					
Yes (%)	51.3	59.1	62.0	60.5	57.2
Ever use of menopausal hormone therapy ‡					
Yes (%)	19.3	23.8	25.0	25.7	27.4
Menopausal status ‡					
Postmenopausal (%)	42.9	41.3	41.3	42.3	43.6

Mean and standard deviation (in parenthesis) for continuous variables, or percentages for categorical variables (‡).

3.1.4 Dietary vitamin D and risk of all-cause mortality

Dietary vitamin D intake was not associated with all-cause mortality risk in the multivariable model (Q5 vs. Q1, HR 0.96, 95% CI: 0.91-1.01; *P*-trend 0.65) (Table 6). Similar associations were observed for men and women (*P*-interaction 0.65). The dietary vitamin D and all-cause mortality associations did not differ across different strata of smoking status (*P*-interaction 0.31), physical activity index (*P*-interaction 0.07), and BMI (*P*-interaction 0.23). Although, an inverse all-cause mortality association was observed amongst former smokers (Q5 vs. Q1, HR 0.88, 95% CI: 0.79-0.98; *P*-trend 0.04), but not amongst never and current smokers. This significant association was not present in the continuous models (HR per 2.5 µg/day increase, 0.98, 95% CI: 0.96-1.01). There was evidence of significant heterogeneity across participant countries (*P*-heterogeneity 0.007) (Figure 4). Positive associations (significant in France and Germany) were observed in half of the participant countries. A similar null association in the sexes combined analysis was observed when deaths recorded within the first 5 years of follow-up were excluded (eliminating 5,250 deaths: Q5 vs. Q1, HR 0.96, 95% CI: 0.90-1.02; *P*-trend 0.78).

3.1.5 Dietary vitamin D and risk of cause-specific mortality

During the follow-up period, cause-specific mortality end-points analysed included: cancer (n=10,157), circulatory diseases (n=5,083), respiratory diseases (n=744), and digestive diseases (n=682).

Cancer mortality
Similar null associations (*P*-interaction 0.61) were observed between dietary vitamin D intake and deaths caused by cancer in men (Q5 vs. Q1, HR 1.01, 95% CI: 0.89-1.16; *P*-trend 0.93) and women (Q5 vs. Q1, HR 0.92, 95% CI: 0.83-1.03; *P*-trend 0.24) (Table 7). Significant heterogeneity was observed across participant countries for the risk of cancer death in the sexes combined analysis (*P*-heterogeneity 0.026); with non-significant risk estimates observed for all countries except a positive association for Germany (HR per 2.5 µg/day, 1.06, 95% CI: 1.02-1.10) and an inverse association for the Netherlands (HR per 2.5 µg/day, 0.74, 95% CI: 0.61-0.90).

Circulatory disease mortality

For deaths caused by circulatory diseases, no associations were observed for men (Q5 vs. Q1, HR 1.05, 95% CI: 0.89-1.25; *P*-trend 0.35) and women (Q5 vs. Q1, HR 1.01, 95% CI: 0.84-1.21; *P*-trend 0.86) (*P*-interaction 0.61) (Table 7). Differences in associations were observed across countries (*P*-heterogeneity 0.03), with non-significant risk estimates observed for all countries except Denmark (HR per 2.5 µg/day, 0.90, 95% CI: 0.82-0.99), Germany (HR per 2.5 µg/day, 1.06, 95% CI: 1.01-1.12), and the Netherlands (HR per 2.5 µg/day, 1.34, 95% CI: 1.06-1.71).

Respiratory disease and digestive disease mortality

Null associations were also observed for deaths caused by respiratory diseases in men (Q5 vs. Q1, HR 1.09, 95% CI: 0.67-1.80; *P*-trend 0.73) and women (Q5 vs. Q1, HR 1.02, 95% CI: 0.66-1.57; *P*-trend 0.38) (*P*-interaction 0.60) (Table 7). Similarly, for digestive disease mortality, no associations were observed among men (Q5 vs. Q1, HR 0.95, 95% CI: 0.60-1.50; *P*-trend 0.85) and women (Q5 vs. Q1, HR 0.99, 95% CI: 0.62-1.60; *P*-trend 0.85) (*P*-interaction 0.09).

3.1.6 Dietary vitamin D and risks of digestive system and non-digestive system cancer mortality

Based on an *a priori* hypothesis that cancers within the digestive system will be the most responsive to vitamin D exposures, deaths were split into digestive system and non-digestive cancers. Overall, no difference was observed in the associations between digestive and non-digestive cancers, with no associations observed in the cancers of both locations (*P*-heterogeneity 0.54) (Table 8). For digestive system cancers, non-significant inverse and positive associations in the categorical models were observed amongst men and women respectively; although this difference was non-significant (*P*-interaction 0.74). Divergent risk estimates were also observed between men and women for non-digestive cancers, with a non-significant positive association observed amongst men and a significant inverse association observed amongst women. This difference by sex was non-significant (*P*-interaction 0.36).

3.1.7   Dietary vitamin D and risks of mortality of individual cancers

Colorectal cancer mortality

Associations between dietary vitamin D intake and risks of death from individual cancers are shown in Table 9. No association was observed for colorectal cancer mortality (Q5 vs. Q1, HR 0.96, 95% CI: 0.75-1.24; *P*-trend 0.53) in analysis of men and women combined. Non-significant results were also observed in separate men and women analyses (*P*-interaction 0.75).

Pancreatic cancer mortality

For pancreatic cancer mortality, a non-significant association was observed for men and women combined (Q5 vs. Q1, HR 1.05, 95% CI: 0.77-1.44; *P*-trend 0.93) and analysed separately (Table 9). Contrasting non-significant inverse and positive associations were observed amongst men and women respectively; although this difference was not significant (*P*-interaction 0.22).

Lung cancer mortality

For lung cancer mortality, no association was observed in the sexes combined model (Table 9). These associations did not differ by sex (*P*-interaction 0.22).

Prostate cancer and breast cancer mortality

Non-significant associations were also observed for prostate cancer mortality (Q5 vs. Q1, HR 0.87, 95% CI: 0.55-1.38; *P*-trend 0.66) and breast cancer mortality (Q5 vs. Q1, HR 1.15, 95% CI: 0.85-1.55; *P*-trend 0.30) (Table 9).

Table 6. Risk (hazard ratios) of death from all causes associated with dietary vitamin D intake

Dietary vitamin D intake range (µg/day)	Sex specific quintile of dietary vitamin D intake						HR (95% CI) per	
	1	2	3	4	5	P-trend	2.5 µg/day	P-interaction
Men	<2.3	2.3-<3.4	3.4-<4.7	4.7-<6.5	≥6.5			
Women	<1.9	1.9-<2.7	2.7-<3.6	3.6-<5.1	≥5.1			
Total all cause mortality								
Overall risk								
N deaths	4,695	4,351	4,915	5,219	5,817			
Person-years	1,123,765	1,159,949	1,163,799	1,158,856	1,145,457			
Basic model - HR (95% CI) *	1.00	0.91 (0.87-0.96)	0.95 (0.91-0.99)	0.92 (0.88-0.97)	0.91 (0.87-0.96)	0.03	1.00 (0.98-1.01)	
Multivariable model - HR (95% CI) †	1.00	0.93 (0.89-0.98)	0.97 (0.93-1.02)	0.96 (0.91-1.00)	0.96 (0.91-1.01)	0.65	1.00 (0.99-1.02)	
Sex								0.65
Men	1.00	0.93 (0.87-1.00)	0.97 (0.91-1.05)	0.97 (0.90-1.05)	0.97 (0.90-1.06)	0.88	1.01 (0.99-1.03)	
Women	1.00	0.93 (0.88-0.99)	0.97 (0.92-1.03)	0.95 (0.89-1.01)	0.95 (0.88-1.02)	0.47	1.00 (0.98-1.02)	
Follow-up								0.85
<5 years	1.00	0.95 (0.86-1.05)	0.97 (0.88-1.07)	0.97 (0.87-1.08)	0.94 (0.83-1.06)	0.44	1.01 (0.98-1.04)	
5-10 years	1.00	0.92 (0.86-0.99)	0.95 (0.89-1.03)	0.95 (0.88-1.03)	0.96 (0.88-1.04)	0.93	1.00 (0.98-1.02)	
≥10 years	1.00	0.93 (0.87-1.00)	0.97 (0.90-1.04)	0.94 (0.87-1.01)	0.95 (0.87-1.04)	0.54	1.00 (0.98-1.03)	
Smoking status								0.31
Never smoked	1.00	0.94 (0.88-1.01)	1.00 (0.93-1.08)	0.98 (0.90-1.06)	1.03 (0.94-1.13)	0.22	1.03 (1.00-1.05)	
Former smoker	1.00	0.91 (0.84-0.99)	0.98 (0.90-1.08)	0.92 (0.84-1.01)	0.88 (0.79-0.98)	0.04	0.98 (0.96-1.01)	
Current smoker	1.00	0.94 (0.86-1.02)	0.93 (0.86-1.01)	0.96 (0.88-1.05)	0.96 (0.88-1.06)	0.90	1.01 (0.99-1.03)	
Body mass index								0.23
<25 kg/m^2	1.00	0.91 (0.85-0.97)	0.96 (0.90-1.02)	0.91 (0.85-0.98)	0.93 (0.85-1.01)	0.29	0.99 (0.97-1.01)	
25.0-29.9 kg/m^2	1.00	0.95 (0.89-1.03)	0.99 (0.92-1.07)	1.01 (0.93-1.09)	1.00 (0.92-1.10)	0.49	1.01 (0.99-1.04)	
≥30 kg/m^2	1.00	0.97 (0.78-1.09)	0.98 (0.87-1.10)	0.95 (0.84-1.08)	0.97 (0.85-1.11)	0.80	1.02 (0.98-1.05)	
Physical activity								0.07
Inactive	1.00	0.94 (0.87-1.02)	1.01 (0.93-1.09)	0.96 (0.88-1.05)	0.99 (0.89-1.09)	0.98	1.00 (0.97-1.03)	
Moderately inactive	1.00	0.92 (0.85-0.99)	0.99 (0.91-1.08)	0.94 (0.87-1.03)	0.94 (0.85-1.04)	0.42	1.00 (0.98-1.03)	
Moderately active	1.00	0.95 (0.85-1.06)	0.93 (0.84-1.04)	0.92 (0.82-1.03)	0.94 (0.83-1.08)	0.66	1.03 (0.99-1.06)	
Active	1.00	0.93 (0.82-1.06)	0.96 (0.85-1.09)	1.03 (0.91-1.17)	1.02 (0.88-1.18)	0.34	1.02 (0.98-1.06)	

* Basic model - Cox regression using total energy intake (continuous), and stratified by age (1-year categories), sex, and centre.

† Multivariable model - Cox regression using total energy intake (continuous), body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none, primary school completed, technical/professional school, secondary school, longer education including university, or not specified), physical activity index (inactive, moderately inactive, moderately active, active, or missing), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown), smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown), ever use of contraceptive pill (yes, no, or unknown), ever use of menopausal hormone therapy (yes, no, or unknown), menopausal status (premenopausal, postmenopausal, perimenopausal/unknown menopausal status, or surgical postmenopausal), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), and intakes of fruits and vegetables (g/day), red and processed meats (g/day), calcium (mg/day), and polyunsaturated fatty acids (g/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Figure 4. Multivariable hazard ratios and 95% confidence intervals of all-cause mortality risk by country, per 2.5 µg/day increase in dietary vitamin D intake

Multivariable model - Cox regression using total energy intake (continuous), body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none, primary school completed, technical/professional school, secondary school, longer education including university, or not specified), physical activity index (inactive, moderately inactive, moderately active, active, or missing), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown), smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown), ever use of contraceptive pill (yes, no, or unknown), ever use of menopausal hormone therapy (yes, no, or unknown), menopausal status (premenopausal, postmenopausal, perimenopausal/unknown menopausal status, or surgical postmenopausal), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), and intakes of fruits and vegetables (g/day), red and processed meats (g/day), calcium (mg/day), and polyunsaturated fatty acids (g/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Table 7. Risk (hazard ratios) of cause-specific deaths associated with dietary vitamin D intake

Causes of Death (ICD-10)	*N* deaths	Sex specific quintile of dietary vitamin D intake					P-trend	HR (95% CI) per 2.5 µg/day increase	Sex *P*-interaction
		1	2	3	4	5			
Men		<2.3	2.3-<3.4	3.4-<4.7	4.7-<6.5	≥6.5			
Women		<1.9	1.9-<2.7	2.7-<3.6	3.6-<5.1	≥5.1			
Cancer (C00-D48)									0.61
Men									
Basic model - HR (95% CI) *	4,307	1.00	0.99 (0.88-1.10)	1.01 (0.90-1.13)	1.00 (0.89-1.13)	0.93 (0.82-1.06)	0.20		
Multivariable model - HR (95% CI) †	4,307	1.00	1.01 (0.91-1.13)	1.05 (0.94-1.18)	1.07 (0.95-1.21)	1.01 (0.89-1.16)	0.93	1.02 (0.99-1.05)	
Women									
Basic model - HR (95% CI) *	5,850	1.00	0.92 (0.84-1.00)	1.01 (0.92-1.10)	0.98 (0.89-1.08)	0.92 (0.83-1.03)	0.27		
Multivariable model - HR (95% CI) †	5,850	1.00	0.92 (0.84-1.01)	1.01 (0.92-1.10)	0.98 (0.89-1.07)	0.92 (0.83-1.03)	0.24	0.99 (0.96-1.01)	
Circulatory diseases (I00-I99)									0.81
Men									
Basic model - HR (95% CI) *	2,780	1.00	1.00 (0.87-1.15)	0.93 (0.81-1.08)	0.96 (0.83-1.12)	1.02 (0.87-1.20)	0.63		
Multivariable model - HR (95% CI) †	2,780	1.00	1.00 (0.87-1.15)	0.94 (0.81-1.09)	0.98 (0.84-1.14)	1.05 (0.89-1.25)	0.35	1.00 (0.96-1.04)	
Women									
Basic model - HR (95% CI) *	2,303	1.00	1.01 (0.88-1.17)	1.00 (0.86-1.16)	1.14 (0.97-1.32)	1.05 (0.88-1.26)	0.46		
Multivariable model - HR (95% CI) †	2,303	1.00	1.01 (0.87-1.17)	0.97 (0.84-1.13)	1.09 (0.93-1.28)	1.01 (0.84-1.21)	0.86	1.00 (0.95-1.06)	
Respiratory diseases (J30-J98)									0.60
Men									
Basic model - HR (95% CI) *	338	1.00	0.86 (0.57-1.30)	0.98 (0.65-1.49)	0.89 (0.58-1.37)	0.86 (0.54-1.39)	0.63		
Multivariable model - HR (95% CI) †	338	1.00	0.96 (0.63-1.46)	1.20 (0.78-1.83)	1.09 (0.70-1.71)	1.09 (0.67-1.80)	0.73	1.04 (0.94-1.16)	
Women									
Basic model - HR (95% CI) *	406	1.00	0.67 (0.47-0.96)	0.83 (0.59-1.17)	0.88 (0.62-1.26)	0.80 (0.53-1.20)	0.78		
Multivariable model - HR (95% CI) †	406	1.00	0.73 (0.51-1.05)	0.90 (0.63-1.29)	1.02 (0.70-1.47)	1.02 (0.66-1.57)	0.38	1.10 (0.95-1.27)	
Digestive diseases (K00-K93)									0.09
Men									
Basic model - HR (95% CI) *	350	1.00	0.78 (0.55-1.12)	0.72 (0.50-1.05)	0.75 (0.51-1.10)	0.61 (0.40-0.95)	0.06		
Multivariable model - HR (95% CI) †	350	1.00	0.86 (0.60-1.24)	0.91 (0.62-1.33)	1.02 (0.69-1.51)	0.95 (0.60-1.50)	0.85	1.02 (0.91-1.14)	
Women									
Basic model - HR (95% CI) *	332	1.00	1.04 (0.72-1.51)	1.09 (0.75-1.58)	1.02 (0.69-1.53)	1.01 (0.64-1.59)	0.92		
Multivariable model - HR (95% CI) †	332	1.00	1.04 (0.71-1.51)	1.07 (0.73-1.56)	0.99 (0.66-1.50)	0.99 (0.62-1.60)	0.85	0.92 (0.77-1.10)	
External causes (S00-Y98)									0.14
Men									
Basic model - HR (95% CI) *	575	1.00	0.73 (0.54-0.99)	1.05 (0.78-1.41)	0.92 (0.68-1.26)	0.82 (0.58-1.16)	0.46		
Multivariable model - HR (95% CI) †	575	1.00	0.76 (0.55-1.03)	1.10 (0.82-1.49)	1.00 (0.73-1.37)	0.91 (0.64-1.29)	0.91	0.97 (0.89-1.06)	
Women									
Basic model - HR (95% CI) *	473	1.00	0.95 (0.71-1.27)	0.89 (0.66-1.21)	0.73 (0.52-1.02)	1.16 (0.81-1.66)	0.38		
Multivariable model - HR (95% CI) †	473	1.00	0.98 (0.73-1.31)	0.93 (0.69-1.27)	0.75 (0.54-1.06)	1.21 (0.84-1.74)	0.30	1.09 (0.99-1.20)	

* Basic model - Cox regression using total energy intake (continuous), and stratified by age (1-year categories), sex, and centre.

† Multivariable model - Cox regression using total energy intake (continuous), body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none, primary school completed, technical/professional school, secondary school; longer education including university, or not specified), physical activity index (inactive, moderately inactive, moderately active, active, or missing), smoking status and intensity (never; current , 1-15 cigarettes per day; current,

16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown), smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown), ever use of contraceptive pill (yes, no, or unknown), ever use of menopausal hormone therapy (yes, no, or unknown), menopausal status (premenopausal, postmenopausal, perimenopausal/unknown menopausal status, or surgical postmenopausal), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), and intakes of fruits and vegetables (g/day), red and processed meats (g/day), calcium (mg/day), and polyunsaturated fatty acids (g/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Table 8. Risk (hazard ratios) of digestive system and non-digestive system deaths associated with dietary vitamin D intake

		Sex specific quintile of dietary vitamin D intake						
Dietary vitamin D intake range (μg/day)		1	2	3	4	5		HR (95% CI) per
Men		<2.3	2.3-<3.4	3.4-<4.7	4.7-<6.5	≥6.5		
Women	N deaths	<1.9	1.9-<2.7	2.7-<3.6	3.6-<5.1	≥5.1	P-trend	2.5 μg/day increase
Digestive system cancers								
Overall risk								
Basic model - HR (95% CI) *	2,674	1.00	0.96 (0.83-1.10)	1.08 (0.94-1.24)	1.10 (0.95-1.27)	0.97 (0.83-1.15)	0.74	
Multivariate model - HR (95% CI) †	2,674	1.00	0.97 (0.84-1.11)	1.09 (0.95-1.25)	1.11 (0.96-1.29)	1.00 (0.85-1.18)	0.95	0.99 (0.95-1.03)
Sex								
Men	1,276	1.00	0.90 (0.74-1.10)	1.00 (0.82-1.23)	1.10 (0.89-1.36)	0.90 (0.71-1.15)	0.66	1.00 (0.94-1.06)
Women	1,398	1.00	1.02 (0.84-1.23)	1.17 (0.97-1.42)	1.12 (0.92-1.37)	1.09 (0.87-1.37)	0.62	0.98 (0.92-1.04)
Non-digestive system cancers								
Overall risk								
Basic model - HR (95% CI) *	7,483	1.00	0.94 (0.86-1.01)	0.97 (0.90-1.06)	0.94 (0.87-1.03)	0.90 (0.82-0.99)	0.04	
Multivariate model - HR (95% CI) †	7,483	1.00	0.95 (0.88-1.03)	1.00 (0.92-1.08)	0.97 (0.89-1.06)	0.94 (0.85-1.04)	0.28	1.00 (0.98-1.03)
Sex								
Men	3,031	1.00	1.07 (0.93-1.22)	1.07 (0.93-1.23)	1.06 (0.92-1.22)	1.07 (0.91-1.25)	0.69	1.02 (0.99-1.06)
Women	4,452	1.00	0.90 (0.81-0.99)	0.96 (0.87-1.07)	0.94 (0.84-1.04)	0.88 (0.77-0.99)	0.10	0.99 (0.96-1.02)

* Basic model - Cox regression using total energy intake (continuous), and stratified by age (1-year categories), sex, and centre.

† Multivariable model - Cox regression using total energy intake (continuous), body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none, primary school completed, technical/professional school, secondary school, longer education including university, or not specified), physical activity index (inactive, moderately inactive, moderately active, active, or missing), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown), smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown), ever use of contraceptive pill (yes, no, or unknown), ever use of menopausal hormone therapy (yes, no, or unknown), menopausal status (premenopausal, postmenopausal, perimenopausal/unknown menopausal status, or surgical postmenopausal), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), and intakes of fruits and vegetables (g/day), red and processed meats (g/day), calcium (mg/day), and polyunsaturated fatty acids (g/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Digestive system cancers were oesophagus (C15), stomach (C16), colon and rectum (C18, C19, C20), anus and anal canal (C21), liver and intrahepatic bile acids (C22), gallbladder (C23), other and unspecified parts of biliary tract (C24), pancreas (C25), and other ill-defined digestive organs (C26). Non-digestive system cancers (all other cancers).

Table 9. Risk (hazard ratios) of cancer deaths associated with dietary vitamin D intake

		Sex specific quintile of dietary vitamin D intake							
Dietary vitamin D intake range (µg/day)		**1**	**2**	**3**	**4**	**5**			
Men		**<2.3**	**2.3-<3.4**	**3.4-<4.7**	**4.7-<6.5**	**≥6.5**		**HR (95% CI) per**	
Women	***N* deaths**	**<1.9**	**1.9-<2.7**	**2.7-<3.6**	**3.6-<5.1**	**≥5.1**	***P*-trend**	**2.5 µg/day increase**	**Sex *P*-interaction**
Colorectal cancer (C18-C20)									
Overall risk									
Basic model - HR (95% CI) *	1,145	1.00	1.02 (0.83-1.26)	1.11 (0.90-1.37)	1.10 (0.88-1.37)	0.96 (0.75-1.23)	0.52		
Multivariate model - HR (95% CI) †¥	1,145	1.00	1.02 (0.82-1.25)	1.10 (0.89-1.36)	1.09 (0.88-1.37)	0.96 (0.75-1.24)	0.53	0.96 (0.90-1.02)	
Sex									0.75
Men	497	1.00	0.92 (0.67-1.27)	0.90 (0.65-1.25)	1.04 (0.74-1.46)	0.82 (0.56-1.21)	0.46	0.97 (0.88-1.07)	
Women	648	1.00	1.08 (0.82-1.44)	1.28 (0.97-1.70)	1.13 (0.84-1.52)	1.08 (0.77-1.52)	0.98	0.95 (0.87-1.03)	
Pancreatic cancer (C25)									
Overall risk									
Basic model - HR (95% CI) *	758	1.00	1.02 (0.79-1.33)	1.17 (0.90-1.52)	1.16 (0.88-1.53)	1.05 (0.77-1.43)	0.93		
Multivariate model - HR (95% CI) †	758	1.00	1.02 (0.78-1.33)	1.16 (0.89-1.52)	1.16 (0.88-1.53)	1.05 (0.77-1.44)	0.93	1.00 (0.92-1.08)	
Sex									0.22
Men	339	1.00	0.99 (0.67-1.47)	1.16 (0.78-1.73)	1.03 (0.68-1.58)	0.89 (0.55-1.44)	0.47	0.95 (0.84-1.07)	
Women	419	1.00	1.04 (0.73-1.48)	1.16 (0.81-1.66)	1.26 (0.87-1.82)	1.20 (0.79-1.84)	0.41	1.04 (0.94-1.16)	
Lung cancer (C34)									
Overall risk									
Basic model - HR (95% CI) *	1,919	1.00	0.96 (0.82-1.13)	1.00 (0.85-1.18)	0.96 (0.81-1.14)	0.90 (0.74-1.09)	0.25		
Multivariate model - HR (95% CI) †	1,919	1.00	1.00 (0.85-1.18)	1.03 (0.87-1.22)	1.02 (0.86-1.22)	1.01 (0.83-1.23)	0.99	1.04 (0.99-1.08)	
Sex									0.40
Men	1,042	1.00	1.12 (0.89-1.40)	1.25 (0.98-1.58)	1.12 (0.87-1.44)	1.19 (0.90-1.57)	0.46	1.06 (1.01-1.10)	
Women	877	1.00	0.89 (0.70-1.13)	0.85 (0.67-1.08)	0.95 (0.74-1.21)	0.87 (0.65-1.16)	0.62	1.01 (0.95-1.08)	
Prostate cancer (C61)									
Overall risk									
Basic model - HR (95% CI) *	392	1.00	0.90 (0.61-1.35)	0.90 (0.60-1.34)	0.81 (0.53-1.22)	0.85 (0.54-1.33)	0.55		
Multivariate model - HR (95% CI) †	392	1.00	0.91 (0.61-1.35)	0.91 (0.61-1.35)	0.82 (0.54-1.25)	0.87 (0.55-1.38)	0.66	0.96 (0.86-1.08)	
Breast cancer (C50)									
Overall risk									
Basic model - HR (95% CI) *	741	1.00	0.92 (0.72-1.18)	1.11 (0.87-1.43)	0.98 (0.75-1.27)	1.11 (0.83-1.49)	0.37		
Multivariate model - HR (95% CI) †Φ	741	1.00	0.93 (0.72-1.19)	1.13 (0.88-1.46)	1.00 (0.76-1.31)	1.15 (0.85-1.55)	0.30	1.07 (1.00-1.13)	

* Basic model - Cox regression using total energy intake (continuous), and stratified by age (1-year categories), sex, and centre.

† Multivariable model - Cox regression using total energy intake (continuous), body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none, primary school completed, technical/professional school, secondary school, longer education including university, or not specified), physical activity index (inactive, moderately inactive, moderately active, active, or missing), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown), smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown), ever use of contraceptive pill (yes, no, or unknown), ever use of menopausal hormone therapy (yes, no, or unknown), menopausal status (premenopausal, postmenopausal, perimenopausal/unknown menopausal status, or surgical postmenopausal), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), and intakes of fruits and vegetables (g/day), red and processed meats (g/day), calcium (mg/day), and polyunsaturated fatty acids (g/day) (all continuous), and stratified by age (1-year categories), sex, and centre. Multivariable model – plus adjustment for age at first pregnancy (<21; 21-<30; 30+ years old; no children; or not specified) and age at menarche (<12; 12-<15; 15+ years old; or not specified).

3.2 Derivation and validation of a predictor 25-hydroxyvitamin D score within the European Prospective Investigation into Cancer and Nutrition

3.2.1 Participant characteristics

Characteristics of the participants used in the derivation and validation of the predictor 25(OH)D score are shown in Table 10. The derivation and validation datasets for men and women were similar across all of the characteristics considered.

Table 10. Characteristics of study participants included in the derivation and validation of the predictor circulating 25-hydroxyvitamin D score

Characteristic	Men			Women		
	Sample for the derivation of the predictor score	Sample for the validation of predictor score	P-value §	Sample for the derivation of the predictor score	Sample for the validation of predictor score	P-value §
N	1,982	984		2,107	1,045	
Age at blood collection (years)	59.2 (7.0)	59.3 (7.2)	0.90	55.5 (8.0)	55.0 (8.8)	0.65
Country (%)			0.82			0.60
Denmark	23.8	24.9		22.5	22.6	
France	-	-		6.9	5.7	
Italy	11.1	11.6		17.0	16.9	
Spain	14.2	12.9		9.2	11.3	
United Kingdom	22.2	22.9		16.5	14.8	
The Netherlands	2.4	3.2		14.4	14.6	
Greece	2.5	3.3		2.4	3.0	
Germany	23.9	21.3		11.1	11.1	
Dietary vitamin D intake (µg/day)	4.5 (2.6)	4.4 (2.3)	0.78	3.4 (1.8)	3.3 (1.8)	0.28
Waist circumference (cm)	96.2 (9.9)	96.4 (9.8)	0.87	81.2 (11.0)	81.8 (11.6)	0.67
Physical activity (%)			0.31			0.46
Inactive	25.1	24.0		23.7	25.5	
Moderately inactive	28.7	30.8		35.0	34.2	
Moderately active	23.4	22.6		22.1	20.2	
Active	22.8	22.6		19.3	20.2	
Smoking status and intensity (%)			0.21			0.70
Never smoker	27.4	24.5		52.0	52.0	
Current, 1-15 cigarettes/day	8.1	8.7		13.4	12.7	
Current, 16-25 cigarettes/day	6.0	7.3		5.9	5.6	
Current, 26+ cigarettes/day	2.6	2.4		1.0	1.1	
Former, quit ≤10 years	12.9	13.0		7.1	9.4	
Former, quit 11-20 years	11.7	12.1		7.2	6.8	
Former, quit 20+ years	18.0	18.2		7.9	7.8	
Current, pipe/cigar/ocassional	10.5	11.9		4.0	4.0	
Current/Former, missing	2.9	1.8		1.6	0.8	
Month of blood draw (%)			0.58			0.41
December	6.6	7.2		5.7	6.7	
January	8.8	8.5		7.6	9.2	
February	10.2	9.2		10.3	8.6	
March	11.2	11.7		10.9	11.5	
April	8.3	7.7		9.1	8.4	
May	8.9	9.0		9.1	8.5	
June	8.0	6.5		7.9	9.1	
July	7.6	7.7		7.1	5.9	
August	6.3	5.9		4.8	4.9	
September	7.4	9.0		8.2	7.9	
October	7.9	7.1		8.7	9.1	
November	8.9	10.4		10.6	10.2	
Ever use of menopausal hormone therapy (%)			-			0.54
No	-	-		69.2	67.6	
Yes	-	-		27.3	28.4	
Unknown	-	-		3.5	4.0	
Source study (%)			0.91			0.34
Colorectal cancer	37.7	37.7		34.9	36.5	
Prostate cancer	33.9	33.2		-	-	
Non-Hodgkin lymphomas	28.4	29.1		26.7	24.6	
Breast cancer	-	-		38.4	39.0	

Mean and standard deviation (in parenthesis) unless stated otherwise.

§ *P*-values for differences between the 25(OH)D derivation and validation subsamples were calculated using linear regression (continuous characteristic variables) and chi-square tests (categorical characteristic variables).

3.2.2 Derivation and validation of predictor 25-hydroxyvitamin D scores

Predictor circulating 25-hydroxyvitaminD score derivation

Correlates/determinants of circulating 25(OH)D considered for inclusion in the predictor score were split into the following categories: age at blood collection and timing of blood collection, dietary variables, anthropometric variables, lifestyle variables, reproductive variables, location of residence, and adjustment for different laboratories and assays. The final 25(OH)D predictor scores for men and women explained 34% and 28% of the overall variance in circulating 25(OH)D respectively (Table 11). Correlates/determinants included within the predictor scores were: age at blood collection, dietary vitamin D intake, waist circumference, physical activity index, month of blood collection, country of residence, smoking status and intensity, ever use of menopausal hormone therapy, and source study and batch of serum samples. The rationale for the inclusion of these correlates/determinants (and the exclusion of the other variables considered) is outlined in the section below.

Age at blood collection and timing of blood collection variables

Age at blood collection was a near statistically significant predictor in men and women. Weak divergent associations were observed in men (1 year age increase associated with a 0.003% higher mean 25(OH)D) and women (1 year increase associated with a -0.002% lower mean 25(OH)D level) (Table 11). Despite these weak associations, age at blood collection remained in both predictor scores as it is a known determinant of vitamin D status. However, when the predictor scores were applied to the validation datasets, age was excluded to minimise the risk of statistical over-adjustment; as age is also an important risk factor (and therefore confounder) for chronic disease incidence and mortality, as well as a correlate/determinant of vitamin D status.

Of the timing of blood collection variables considered, only month of blood collection was included in the predictor scores. Month of blood collection explained the most variance in circulating 25(OH)D in the men (16%) and women's (12%) predictor scores. Using December as the reference category, the highest mean circulating 25(OH)D was found amongst those participants who had their blood collected in August in men (41% higher) and August and September in women (both 29% higher) (Table 11 and Figure 5). For both sexes, participants whose blood was collected in March had the lowest mean circulating 25(OH)D levels when compared against the December reference category (16% lower in men and 17% lower in women). Month of blood collection reflects recent exposures

impacting on vitamin D status and is not a factor in determining long-term between person variance; because of this, month of blood collection was excluded when the predictor scores were applied to the validation datasets.

Month of blood collection was selected ahead of season of blood collection as the latter variable explained less variance in circulating 25(OH)D (0.05% less in men and 0.03% less in women). Year of blood collection was not significant (P-value >0.05% in men and women) when included in the predictor scores and as a consequence was excluded.

Dietary variables

Dietary variables considered *a priori* for inclusion in the predictor scores were intakes of vitamin D, calcium and retinol. For dietary vitamin D, amongst both men and women, a 1 μg/day increment in intake was statistically significantly associated with mean 2% higher circulating 25(OH)D levels. In men, 1% of the variance of circulating 25(OH)D levels was explained by dietary vitamin D intakes; whilst negligible variance was explained in women. (Table 11) Dietary calcium intake was not significantly associated with circulating 25(OH)D in men (P-value 0.62) and women (P-value 0.60); in both sexes zero additional variance in circulating 25(OH)D was explained by the addition of dietary calcium to the final predictor scores. Similarly, dietary retinol intake was not associated with circulating 25(OH)D (men: P-value 0.63; women P-value 0.45),and no additional variance in circulating 25(OH)D was explained by its inclusion in the final predictor scores. Of the dietary intake variables considered, only vitamin D was included in the predictor scores.

Anthropometric variables

The anthropometric variables considered for inclusion in the predictor scores as indicators of adiposity were waist circumference, BMI, height, weight, and waist-to-hip ratio. Waist circumference was significantly associated with circulating 25(OH)D in both men and women (P-value <0.001). For both sexes, a 1 cm increase in waist circumference was associated with a mean 1% reduction in circulating 25(OH)D. Waist circumference explained more of the variance in circulating 25(OH)D among women (3%) than in men (1%) (Table 11). BMI was also significantly associated with circulating 25(OH)D when included in the predictor scores instead of waist circumference (P-value <0.001); however, slightly less variance was explained in the predictor scores than when waist circumference was included (0.01% less for men and 0.003% less for women). The close correlation between BMI and waist circumference (Pearson r =0.85 in men; r =0.86 in women) meant that only waist circumference was included in the final predictor scores.

Height (men *P*-value 0.94; women *P*-value 0.06) and weight (men *P*-value 0.26; women *P*-value 0.90) were non-significant when included, in addition to waist circumference, in the final predictor scores; negligible additional variance in circulating 25(OH)D was explained by their inclusion. Similarly, waist-to-hip ratio was non-significant (men *P*-value 0.91; women *P*-value 0.17), and did not explain any additional variance in circulating 25(OH)D when added to the final predictor scores.

In summary, waist circumference was included as the indicator of adiposity as it was of greater statistical significance and/or explained more variance in circulating 25(OH)D than BMI, weight, or waist-to-hip-ratio.

Lifestyle variables

The *a priori* lifestyle variables considered for inclusion were physical activity, smoking status and intensity, educational level, alcohol consumption, and ever use of menopausal hormone therapy. Amongst men and women, increasing physical activity index levels were associated with higher circulating 25(OH)D levels (Table 11). Men, categorised as being physically 'active' had mean 17% higher circulating 25(OH)D levels than those categorised as physically 'inactive'. Amongst women, 'active' participants had a mean circulating 25(OH)D levels 12% higher than those classified as 'inactive'. For both men and women, 1% of the variance in circulating 25(OH)D was explained by the inclusion of physical activity index within the predictor scores. The physical activity index variable was included within the predictor scores to act as a surrogate measure for sun exposure. The physical activity index variable encompasses all occupational and recreational activity.

For smoking status and intensity, amongst both men and women, never smokers had higher circulating 25(OH)D status than current smokers (Table 11). The lowest circulating 25(OH)D levels were observed amongst current smokers of 16-25 cigarettes/day (men: 23% lower mean 25(OH)D compared to never smokers; women: 12% lower compared to never smokers). The highest circulating 25(OH)D levels were observed amongst men and women former smokers, who had quit for 11-20 years (men: 5% higher mean 25(OH)D compared to never smokers; women: 8% higher mean compared to never smokers). Smoking status and intensity explained 2% of circulating 25(OH)D variance in men and 1% in women. Due to smoking behaviour being an important risk factor/confounder in chronic disease/mortality risk models, as well as a significant correlate of circulating 25(OH)D levels, smoking status and

intensity was excluded when the predictor scores were applied to the validation datasets to minimise the risk of statistical over-adjustment.

Education level was not included in the final predictor scores as was non-significant (*P*-value >0.05 in men and women) and explained zero additional variance in circulating 25(OH)D. Alcohol consumption was not significantly associated with circulating 25(OH)D in men (*P*-value 0.81); while amongst women, this association was significant (*P*-value 0.01), but zero additional variance in circulating 25(OH)D was explained by its inclusion. As a consequence alcohol consumption was excluded from both sexes final predictor scores.

Reproductive variables

Women who had previously used menopausal hormone therapy had a mean 5% higher circulating 25(OH)D levels than never users (Table 11). Menopausal status (*P*-value >0.05), ever use of the contraceptive pill (*P*-value >0.05) and whether women had live born children (*P*-value 0.62) were excluded from the final predictor scores as all were non-significantly associated with circulating 25(OH)D in women.

Location of residence

Country of residence explained 1% and 2% of the variance in circulating 25(OH)D amongst men and women respectively. Men from Italy and Germany had the highest and lowest mean circulating 25(OH)D respectively when compared against participants from Denmark (reference category) (Table 11 and Figure 6). Men from Spain had a lower mean circulating 25(OH)D levels than those from the UK and the Netherlands. Amongst women, the highest mean circulating 25(OH)D levels were observed amongst participants from Denmark (reference category); with the lowest observed in participants from Germany (27% lower than Denmark), Greece (22% lower), and Italy (21% lower).

Adjustment for different laboratories and assays

The source study and batch of circulating sample analysis was adjusted for in the final model, but excluded when the predictor score was applied to the validation datasets. Amongst both men and women, inclusion of this adjustment within the predictor scores explained 7% of the overall variances in circulating 25(OH)D (Table 11).

Table 11. Predictors of circulating 25-hydroxyvitamin D included in the scores. Analyses were completed in the derivation datasets of 1,982 men and 2,107 women

Predictor	Men				Women			
	N	Log$_e$ β-coeff.	P-value	R^2	N	Log$_e$ β-coeff.	P-value	R^2
N	1,982	-	-	-	2,107	-	-	-
Age at blood collection (years) ‡	1,982	0.003	0.09	0.001	2,107	-0.002	0.05	0.001
Dietary vitamin D intake (µg/day)	1,982	0.02	<0.001	0.01	2,107	0.02	<0.001	0.00
Waist circumference (cm)	1,982	-0.01	<0.001	0.01	2,107	-0.01	<0.001	0.03
Physical activity index				0.01				0.01
Inactive	497	REF	REF		499	REF	REF	
Moderately inactive	569	0.08	0.001		737	0.09	<0.001	
Moderately active	464	0.11	<0.001		465	0.10	<0.001	
Active	452	0.17	<0.001		406	0.12	<0.001	
Month of blood collection §				0.16				0.12
December	174	REF	REF		120	REF	REF	
January	202	-0.03	0.56		159	-0.05	0.31	
February	221	-0.15	<0.001		216	-0.16	<0.001	
March	165	-0.16	<0.001		230	-0.17	<0.001	
April	176	-0.07	0.12		191	-0.15	0.001	
May	158	0.06	0.23		192	-0.05	0.29	
June	150	0.19	<0.001		167	0.09	0.04	
July	124	0.23	<0.001		150	0.18	<0.001	
August	147	0.41	<0.001		102	0.29	<0.001	
September	157	0.39	<0.001		173	0.29	<0.001	
October	177	0.26	<0.001		184	0.16	<0.001	
November	131	0.07	0.11		223	0.01	0.78	
Country				0.01				0.02
Denmark	472	REF	REF		473	REF	REF	
France	-	-	-		145	-0.16	0.01	
Italy	219	0.05	0.37		359	-0.21	<0.001	
Spain	281	0.01	0.80		193	-0.03	0.54	
United Kingdom	439	0.03	0.56		348	-0.17	0.001	
The Netherlands	48	0.04	0.55		304	-0.12	0.02	
Greece	50	0.04	0.58		51	-0.22	0.004	
Germany	473	-0.10	0.04		234	-0.27	<0.001	
Smoking status and intensity ‡				0.02				0.01
Never smoker	542	REF	REF		1,095	REF	REF	
Current, 1-15 cigarettes/day	161	-0.05	0.14		282	-0.06	0.04	
Current, 16-25 cigarettes/day	118	-0.23	<0.001		124	-0.12	0.01	
Current, 26+ cigarettes/day	52	-0.15	0.06		22	-0.07	0.43	
Former, quit ≤10 years	255	-0.001	0.99		149	0.05	0.18	
Former, quit 11-20 years	232	0.05	0.07		151	0.08	0.03	
Former, quit 20+ years	357	0.01	0.71		166	0.02	0.55	
Current, pipe/cigar/ocassional	208	-0.08	0.03		85	-0.05	0.24	
Current/Former, missing	57	0.09	0.10		33	-0.11	0.09	
Ever use of menopausal hormone therapy				-				0.002
No	-	-	-		1,458	REF	REF	
Yes	-	-	-		576	0.05	0.01	
Unknown	-	-	-		73	0.04	0.46	
Source study and batch of serum samples **			0.07				0.07	
Full model			0.34				0.28	

‡ Age at blood collection and smoking status and intensity were excluded when predictor score was applied to the validation and full cohort datasets, so these variables can be used as confounders (avoiding statistical over-adjustment) when assessing disease risk with the predictor score.

§ Month of blood collection was excluded when predictor score applied to the validation and full cohort datasets, as timing of blood collection reflects recent sun and diet/supplements exposures and is not a factor in determining long-term average between person variation.

** Adjusted for source study and batch of circulating 25(OH)D samples (colorectal cancer, prostate cancer, breast cancer or lymphoma) but excluded when predictor score applied to validation and full cohort datasets.

Figure 5. Geometric mean (95% CI) monthly variation in circulating 25-hydroxyvitamin D in men and women in the European Prospective Investigation into Cancer and Nutrition Adjusted for other correlates/determinants of circulating 25-hydroxyvitamin D detailed in Table 11. Analyses were done in the derivation datasets of 1,982 men and 2,107 women.
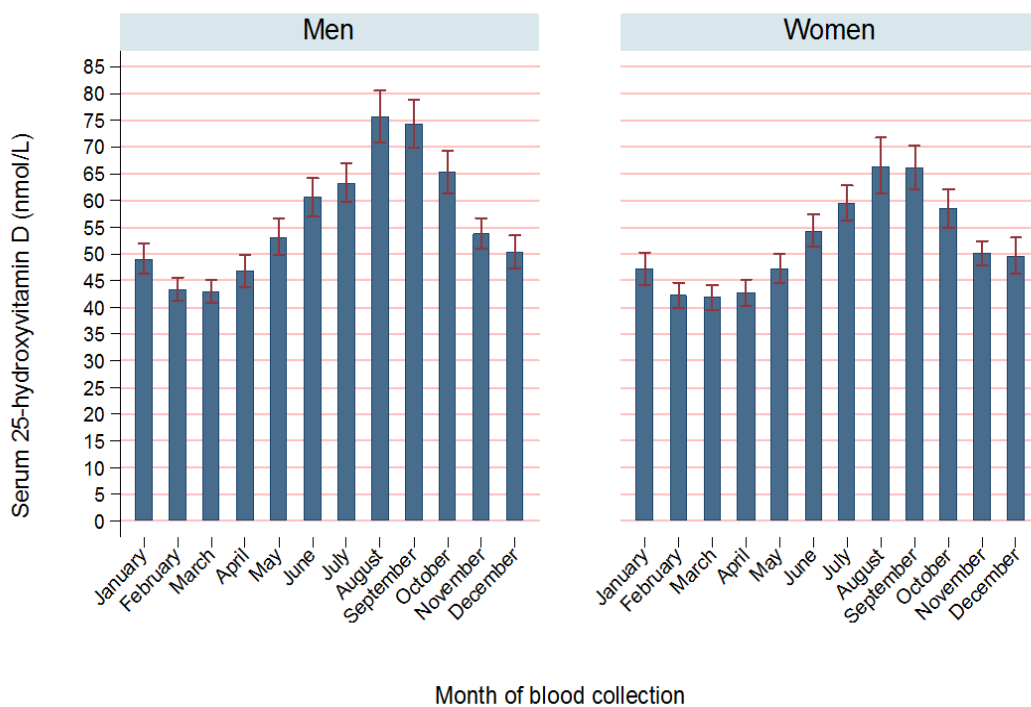
Geometric mean (95% CI) by month of blood collection and sex

Figure 6. Geometric mean (95% CI) country variation in circulating 25-hydroxyvitamin D in men and women in the European Prospective Investigation into Cancer and Nutrition Adjusted for other correlates/determinants of circulating 25-hydroxyvitamin D detailed in Table 11. Analyses were done in the derivation datasets of 1,982 men and 2,107 women.
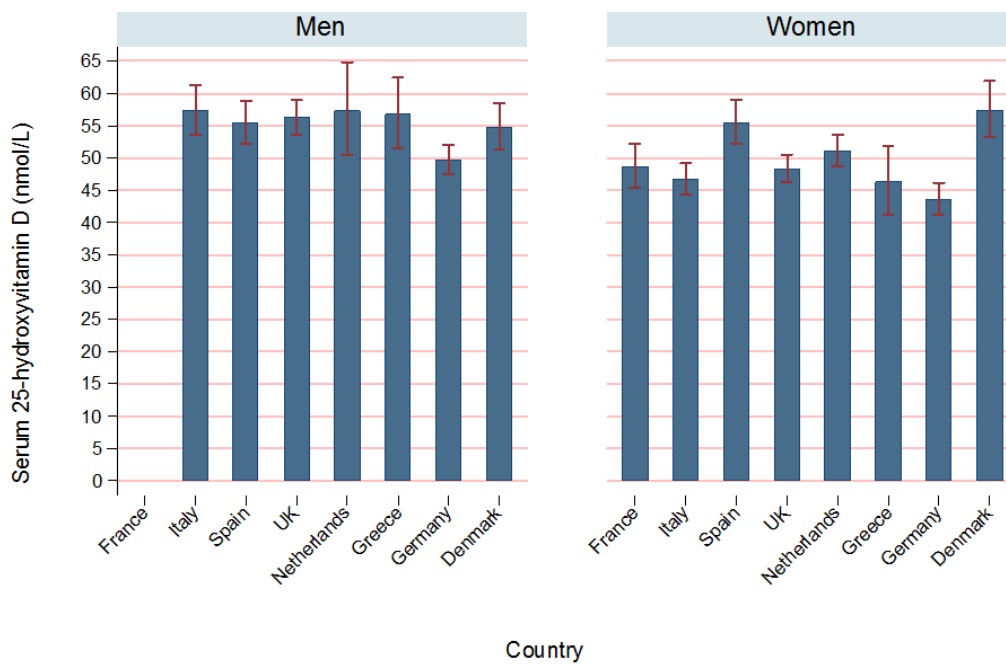
Geometric mean (95% CI) by country and sex

Predictor 25-hydroxyvitamin D scores validation

The separate predictor scores for men and women were then applied to the validation datasets which had actual circulating 25(OH)D measurements for the participants (n=984 in men; and n=1,045 in women). Application of the predictor scores created a predicted 25(OH)D value for each participant.

Correlations between predicted 25-hydroxyvitamin D and actual circulating 25-hydroxyvitamin D

For men, the mean (SD) predicted 25(OH)D level was 46.1 (4.6) nmol/L, compared to an actual circulating 25(OH)D mean of 59.6 (24.6) after adjusting for source study and batch of

100

laboratory analysis. The Pearson and Spearman correlation coefficients between these variables were identical: 0.20 (95% CI: 0.14-0.26). When the actual circulating 25(OH)D measurements were further adjusted for month of blood collection and age of blood collection (mean actual 25(OH)D: 58.7 (22.3) nmol/L), the correlations increased slightly (Pearson: 0.21, 95% CI: 0.15-0.27; Spearman: 0.21, 95% CI: 0.14-0.26). Amongst women, the mean (SD) predicted 25(OH)D level was 55.2 (8.4) nmol/L, compared to an actual circulating 25(OH)D mean of 52.9 (21.9) nmol/L after adjusting for source study and batch of laboratory analysis. The Pearson and Spearman correlation coefficients were 0.19 (95% CI: 0.13-0.25) and 0.20 (95% CI: 0.14-0.26) respectively. Similarly to men, further adjusting the actual 25(OH)D levels for month of blood collection and age at blood collection slightly increased these correlations (mean actual 25(OH)D: 52.1 (19.5) nmol/L; Pearson: 0.20, 95% CI: 0.14-0.26; Spearman: 0.21, 95% CI: 0.16-0.27).

Actual 25-hydroxyvitaminD measurements by quantile of predicted 25-hydroxyvitamin D

For both men and women actual measurements generally increased with increasing quintile and decile of predicted 25(OH)D levels (Figure 7). The differences between the extreme predicted quintiles of mean actual 25(OH)D was 18.3 nmol/L in men and 13.2 nmol/L in women, with incremental increases in actual 25(OH)D generally observed in the intermediate categories (*P*-trends <0.0001) (Figure 7A). Across deciles of predicted 25(OH)D, linear increases in mean actual 25(OH)D measurements were observed in men and women (*P*-trends <0.0001). The differences in mean actual 25(OH)D between the extreme predicted 25(OH)D deciles was 23.3 nmol/L in men and 15.2 nmol/L in women. For both men and women, higher mean actual 25(OH)D values were generally yielded across incremental increases in predicted 25(OH)D deciles (Figure 7B).

Figure 7. Mean actual circulating 25-hydroxyvitamin D level by: (A) quintile; (B) decile of predicted 25-hydroxyvitamin D score in the European Prospective Investigation into Cancer and Nutrition

Analyses were done in the validation datasets of 984 men and 1,045 women.

(A)
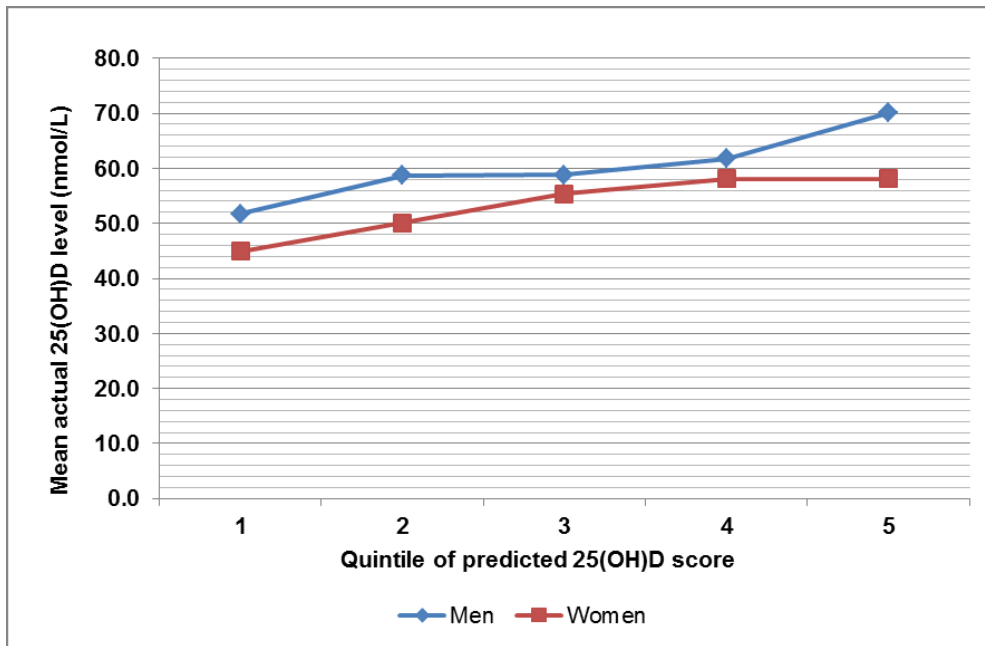
(B)
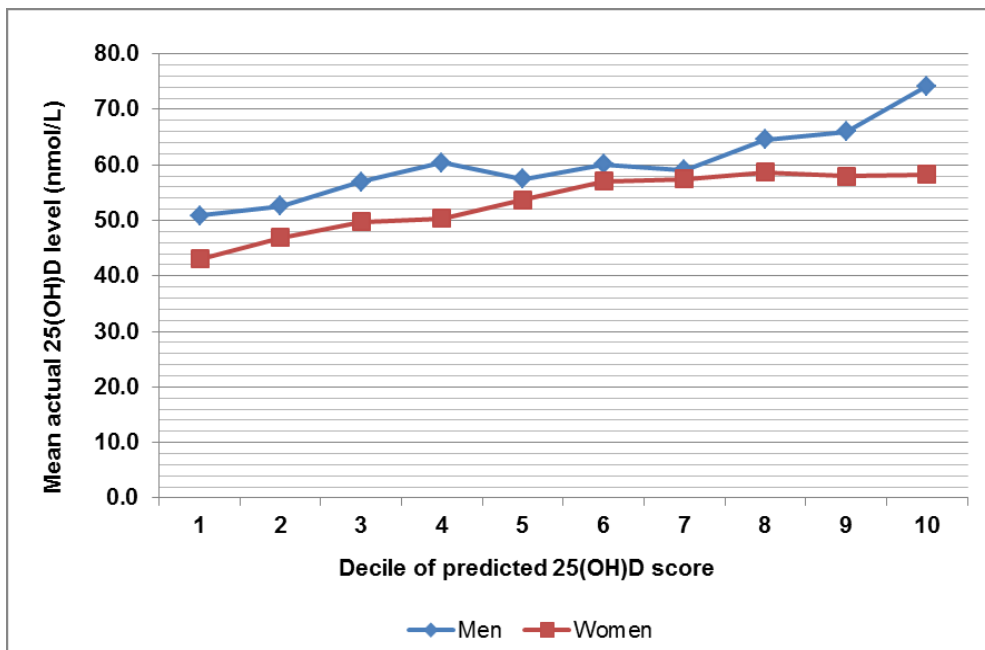
Cross-classifications of participants by predicted and actual circulating 25-hydroxyvitamin D categories

How well the predictor scores categorised participants according to clinically relevant levels of circulating 25(OH)D, tertiles, and quintiles is shown in Figure 8. Figure 8A shows the agreement between participants' predicted and actual 25(OH)D using three clinically relevant categories: <50; 50 to <75; and ≥75 nmol/L. Of the 384 men classified as having a deficiency in vitamin D status (<50 nmol/L) using their actual 25(OH)D measurements, 85.4% (n=328) of them were classified into the corresponding predicted 25(OH)D category. In the mid-range exposure category, 19.7% (n=74) of men were similarly classified into matching predicted and actual 25(OH)D categories. Of the 225 men in the ≥75 nmol/L actual 25(OH)D category, none were similarly classified into the equivalent predicted 25(OH)D category. Amongst the 525 women, classified in the <50 nmol/L actual circulating 25(OH)D category, 35.4% were classified into the corresponding predicted 25(OH)D category (Figure 8A). Of the 354 women classified in the actual 25(OH)D 50 to <75 nmol/L category, 77.7% (n=275) were classified in the analogous predicted 25(OH)D category. Finally, of the 166 women classified in the ≥75 nmol/L actual 25(OH)D category, just 2.4% (n=4) were classified into the equivalent predicted 25(OH)D category.

In epidemiological studies assessing the circulating 25(OH)D-disease relationships five pre-defined clinically relevant categories are used when analysing the associations: <25; 25 to <50; 50 to <75; 75 to <100; ≥100 nmol/L (67;176). The outermost categories reflect the extremes of severe deficiency (<25 nmol/L) through to high exposure status (≥100 nmol/L). Using the predictor scores, such extremes of exposure within EPIC could not be identified as participants were solely classified into the 25 to <50 and 50 to <75 nmol/L categories.

Another approach to analyse relationships in epidemiological studies is to categorise the exposures into quantiles. In Figure 8B, the agreement between participants classified into predicted and actual circulating 25(OH)D tertiles is shown. Of the 328 men and 349 women classified in the lowest exposure tertile of actual circulating 25(OH)D, 41.2% and 45.6% were similarly classified into the parallel lowest tertile of predicted 25(OH)D. One-third of participants (32.9% in men and 33.9% in women) were jointly classified in the same predicted and actual circulating 25(OH)D second tertile. Agreement in participant's classification between predicted tertile 3 to actual circulating 25(OH)D tertile 3 was 42.1% in men and 41.4% in women. Overall, across all tertiles of actual 25(OH)D, 39% and 40% of men and women were classified into the corresponding predicted 25(OH)D tertile.

Figure 8C shows the proportion of participants classified by quintiles of actual and predicted 25(OH)D. Of men classified in the first actual circulating 25(OH)D quintile: 32.2% were classified into predicted 25(OH)D quintile 1; 51.3% were classified into predicted 25(OH)D quintile 1 or 2; and 10.6% were classified into predicted 25(OH)D quintile 5. Of those men classified into the highest actual circulating 25(OH)D quintile: 30.8% were classified into the corresponding predicted 25(OH)D quintile; 51.3% were classified either in quintiles 4 and 5 of predicted 25(OH)D; and 11.3% were classified into the lowest predicted 25(OH)D quintile. Of women classified in the lowest quintile of actual circulating 25(OH)D: 31.1% were classified into the equivalent predicted 25(OH)D category; 52.6% were classified into predicted quintiles 1 and 2; and 15.8% were classified into quintile 5 of predicted 25(OH)D. Amongst women who were classified into the highest actual circulating 25(OH)D quintile: 25.4% were classified into the analogous predicted 25(OH)D category; 52.6% were classified into predicted 25(OH)D quintiles 4 and 5; and 11.5% were classified into the lowest predicted 25(OH)D quintile. For men and women, 46.3% and 45.8% were respectively classified into equivalent or parallel quintiles of actual and predicted 25(OH)D (Figure 9).

Assessment of colorectal cancer risk in the nested case-control dataset

The next stage in the validation process was to assess the risks of colorectal cancer incidence using the predictor scores and then to compare these results against previously published studies within EPIC. A nested case-control study carried out by Jenab *et al.,* (67) previously reported 38% reduced colorectal cancer incidence risk (95% CI: 0.47-0.81) when the highest and lowest quintiles of circulating 25(OH)D were compared (Table 12). This association was linear (*P*-trend <0.001) and standardised for month of blood collection. In the continuous model, a 4% lower colorectal cancer risk (95% CI: 0.94-0.98) was observed per higher 5.9 nmol/L (100 IU/L) increase in circulating 25(OH)D. When the predictor scores (re-derived with the samples sourced from the colorectal cancer nested case-control study excluded) were applied to the nested case-control dataset significant inverse associations were also observed (Table 12). These associations were stronger than when actual circulating 25(OH)D measurements were used in the models. In the categorical multivariable model, a 50% reduced colorectal cancer risk was observed amongst participants in the highest predicted 25(OH)D quintile when compared against those in the lowest quintile; this association was linear (*P*-trend 0.03). In the continuous model, a 20% lower colorectal cancer risk (95% CI: 0.68-0.93) was observed per increment of 5.9 nmol/L (100 IU/L). Additional adjustment for physical activity index may be deemed as statistical over-adjustment as was an important component of the predictor scores, serving as a proxy indicator of sun exposures; however, when adjusted for in the multivariable models similar

associations were observed. When the multivariable model was not adjusted for BMI (due to its close correlation with waist circumference (r =78) – the marker of adiposity used within the predictor scores - a slightly stronger significant inverse association was observed.

Assessment of colorectal cancer risk in the full European Prospective Investigation into Cancer and Nutrition cohort

When the predictor scores were applied to the full EPIC cohort - minus the participants whose samples were used in the derivation of the predictor scores - similar inverse results were also observed to the Jenab *et al.,* (67) analysis. With men and women analysed jointly, a 25% reduced risk of colorectal cancer incidence was observed in the multivariable model (Table 13). This association was linear (*P*-trend 0.009); and in the continuous model, a 6% lower colorectal cancer incidence risk was observed per 5.9 nmol/L higher predicted 25(OH)D. This statistical model adjusted for identical covariates as the Jenab *et al.,* (67) models, except for the timing of blood collection (not a relevant factor in determining long-term between person variation) and physical activity. However, for completeness, the results with this additional adjustment are presented. In the continuous models, the risk estimates were attenuated with statistical significance lost (HR per 5.9 nmol/L increment in predicted 25(OH)D, 0.97, 95% CI: 0.89-1.05). Conversely, when the multivariable model was not adjusted for BMI, stronger inverse associations were observed.

Assessment of prostate cancer risk in the full European Prospective Investigation into Cancer and Nutrition cohort

The association between circulating 25(OH)D and prostate cancer incidence has previously been investigated within EPIC in a nested case-control study by Travis *et al.,* (97). In this analysis, a non-significant suggestive positive association was observed in the multivariable models standardised by month of blood collection (Q5 vs. Q1, OR 1.28, 95% CI: 0.88-1.88). This association was non-linear (*P*-trend 0.19) and in the continuous models, a non-significant 5% higher prostate cancer risk was observed per 25 nmol/L increment in circulating 25(OH)D. After the predictor scores were applied to the full EPIC cohort, a non-significant weak positive association was observed in the multivariable model (Q5 vs. Q1, OR 1.08, 95% CI: 0.92-1.26) (Table 14). In the continuous model, a 7% greater prostate cancer risk was observed per 25 nmol/L higher circulating 25(OH)D. Identical covariate adjustments were made as in the Travis *et al.,* (97) analysis; except for the variables reflecting the timing of blood collection and, once more, physical activity. When the multivariable models were additionally adjusted for physical activity, and not adjusted for BMI, similar risk estimates were observed.

Figure 8. The percentage of participants who were classified into predicted circulating 25-hyrdroxyvitamin D by their actual circulating 25(OH)D categories

Analyses were done in the validation datasets of 984 men and 1,045 women. Categories: (A) Three clinically defined categories (<50; 50 to <75; ≥75 nmol/L); (B) tertiles; and (C) quintiles. Red box denotes the percentage of participants whose predicted category corresponds to their actual circulating 25(OH)D category.

Men

(A)

Women

Men

(B)

Women

Men   **Women**

(C)
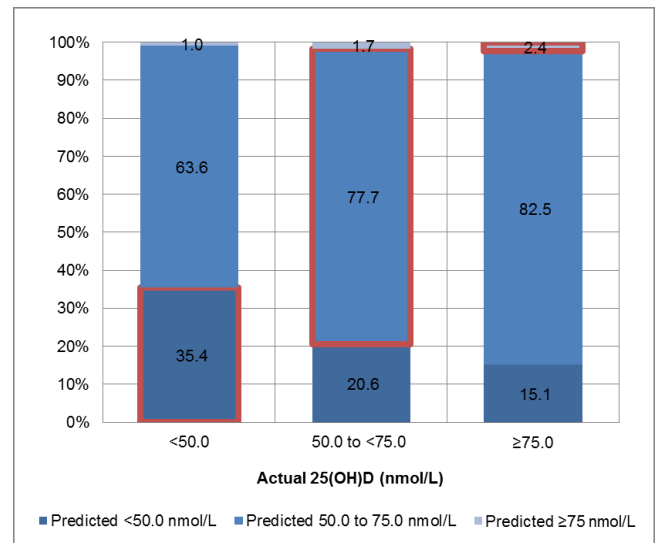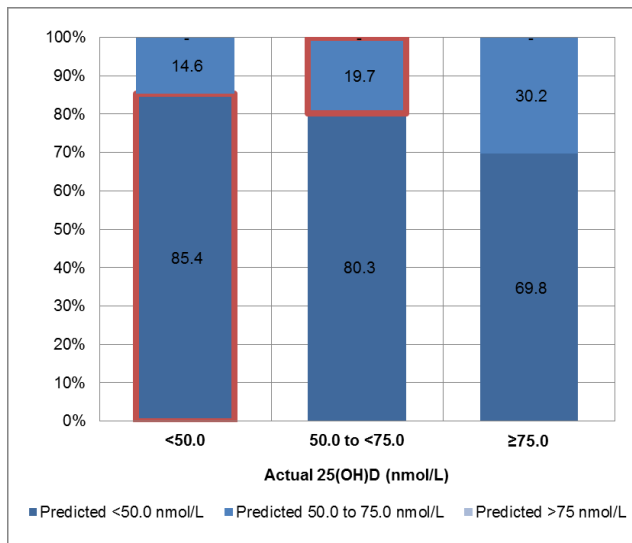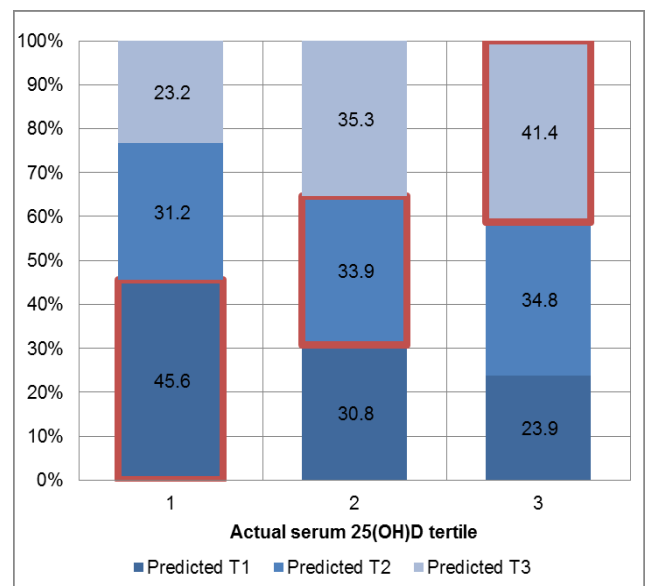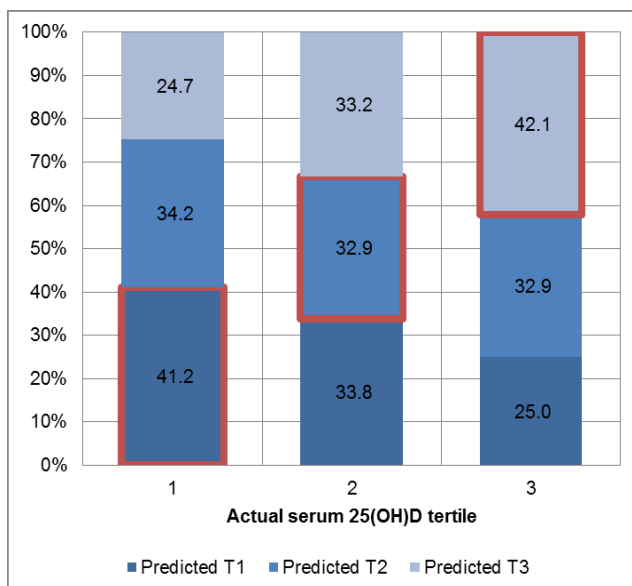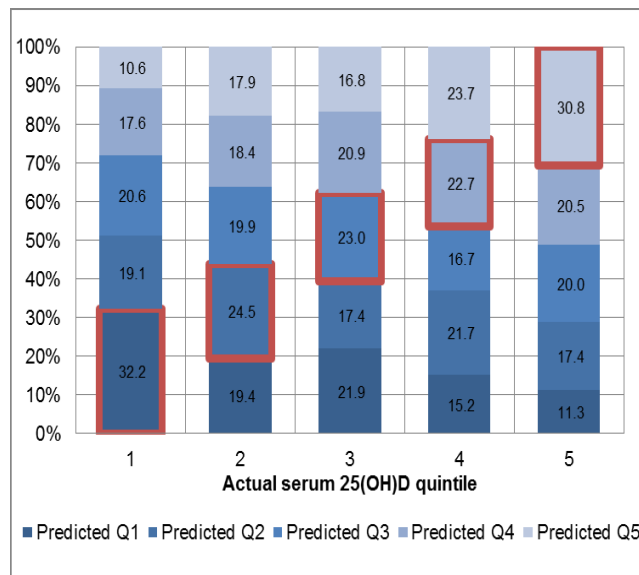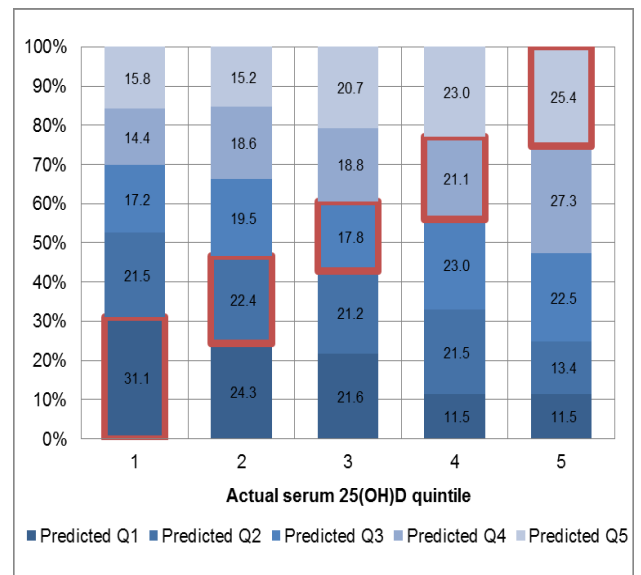
Figure 9. The percentage of participants (by sex) classified by quintiles of actual and predicted circulating 25-hydroxyvitamin D

Analyses were done in the validation datasets of 984 men and 1,045 women.
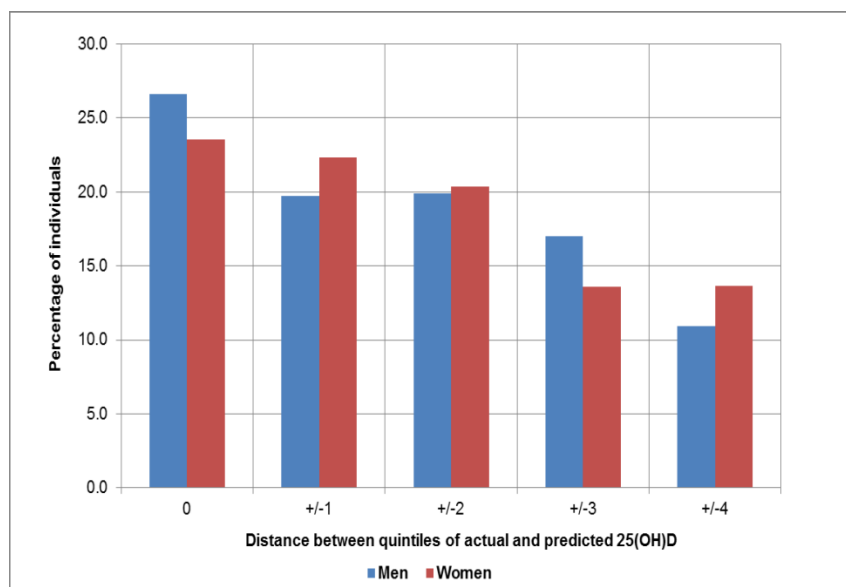
Table 12. Assessment of colorectal cancer incidence risk using the predictor 25-hydroxyvitamin D score in the colorectal nested case-control dataset, compared with a previous published analysis within the European Prospective Investigation into Cancer and Nutrition (67)

	Standardised by month						Non-standardised
	Q1	Q2	Q3	Q4	Q5		
Jenab *et al.*, 2010	≤40.2	>40.2-≤53	>53-≤64.2	>64.2-≤79.8	>79.8	*P*-trend	RR per 100 IU/L (5.9 nmol/L)
Matching variables only †	1.00	0.82 (0.65-1.05)	0.78 (0.60-1.00)	0.63 (0.49-0.82)	0.61 (0.47-0.79)	<0.001	
Multivariable model ɸ	1.00	0.80 (0.62-1.03)	0.78 (0.60-1.01)	0.62 (0.48-0.82)	0.62 (0.47-0.81)	<0.001	0.96 (0.94-0.98)
Predicted 25(OH)D score	Q1	Q2	Q3	Q4	Q5	*P*-trend	RR per 100 IU/L (5.9 nmol/L)
Excluding colorectal cancer serum samples	≤42.0	>42.0-≤46.1	>46.1-≤52.8	>52.8-≤61.5	>61.5		
Matching variables only †	1.00	0.66 (0.49-0.89)	0.59 (0.42-0.83)	0.55 (0.34-0.88)	0.42 (0.24-0.72)	0.001	
Multivariable model ɸ minus BMI	1.00	0.67 (0.49-0.91)	0.60 (0.42-0.86)	0.55 (0.34-0.89)	0.41 (0.24-0.72)	0.001	0.79 (0.70-0.89)
Multivariable model ɸ	1.00	0.70 (0.51-0.96)	0.65 (0.44-0.96)	0.63 (0.36-1.09)	0.50 (0.26-0.97)	0.03	0.80 (0.68-0.93)
Multivariable model ɸ plus physical activity adjustment	1.00	0.72 (0.51-1.01)	0.68 (0.45-1.04)	0.64 (0.35-1.18)	0.51 (0.24-1.07)	0.07	0.79 (0.66-0.95)

† Matched by centre, sex, age at recruitment, and time of day of blood collection.

ɸ Multivariable model - adjusted for BMI, smoking, alcohol intake, education, physical activity, total energy intake, and intakes of total fruits and vegetables, meats/meat products, and alcohol.

Table 13. Assessment of colorectal cancer incidence risk within the full European Prospective Investigation into Cancer and Nutrition cohort using the predictor 25-hydroxyvitamin D score

Analyses were done in the full cohort containing 112,957 men and 217,467 women.

Predicted 25(OH)D score	Q1	Q2	Q3	Q4	Q5		
Men	<42.8	42.8-45.6	45.7-48.1	48.2-50.9	≥51.0		
Women	<48.0	48.0-52.4	52.5-56.1	56.2-60.3	≥60.4	*P*-trend	HR per 100 IU/L (5.9 nmol/L)
Both sexes							
N cases	517	426	396	392	381		
Basic model §	1.00	0.78 (0.68-0.89)	0.74 (0.64-0.86)	0.76 (0.66-0.89)	0.65 (0.55-0.76)	<0.001	0.89 (0.84-0.93)
Multivariable model ¶ minus BMI	1.00	0.78 (0.68-0.90)	0.75 (0.65-0.87)	0.78 (0.67-0.91)	0.66 (0.56-0.79)	<0.001	0.89 (0.85-0.94)
Multivariable model ¶	1.00	0.82 (0.71-0.95)	0.80 (0.69-0.94)	0.85 (0.72-1.01)	0.75 (0.62-0.91)	0.009	0.94 (0.88-0.99)
Multivariable model ¶ plus physical activity	1.00	0.85 (0.73-0.98)	0.85 (0.71-1.01)	0.90 (0.74-1.10)	0.80 (0.63-1.01)	0.12	0.97 (0.89-1.05)
By sex							
Men							
N cases	251	187	178	182	145		
Basic model §	1.00	0.72 (0.59-0.87)	0.67 (0.55-0.83)	0.73 (0.59-0.91)	0.62 (0.49-0.79)	0.001	0.83 (0.76-0.92)
Multivariable model ¶ minus BMI	1.00	0.73 (0.60-0.89)	0.69 (0.56-0.86)	0.76 (0.61-0.95)	0.67 (0.53-0.84)	0.002	0.86 (0.78-0.95)
Multivariable model ¶	1.00	0.79 (0.64-0.97)	0.77 (0.62-0.96)	0.88 (0.69-1.11)	0.79 (0.61-1.03)	0.29	0.94 (0.84-1.05)
Multivariable model ¶ plus physical activity	1.00	0.82 (0.66-1.03)	0.84 (0.64-1.09)	0.96 (0.70-1.30)	0.85 (0.59-1.22)	0.85	0.98 (0.83-1.16)
Women							
N cases	266	239	218	210	236		
Basic model §	1.00	0.84 (0.69-1.01)	0.81 (0.66-0.99)	0.79 (0.64-0.98)	0.66 (0.52-0.85)	0.003	0.91 (0.85-0.97)
Multivariable model ¶ minus BMI	1.00	0.84 (0.69-1.01)	0.80 (0.66-0.98)	0.79 (0.64-0.98)	0.66 (0.52-0.85)	0.002	0.91 (0.85-0.97)
Multivariable model ¶	1.00	0.85 (0.70-1.04)	0.83 (0.67-1.04)	0.82 (0.64-1.05)	0.70 (0.52-0.95)	0.048	0.92 (0.85-0.99)
Multivariable model ¶ plus physical activity	1.00	0.87 (0.71-1.07)	0.86 (0.68-1.10)	0.86 (0.64-1.14)	0.75 (0.53-1.05)	0.18	0.94 (0.85-1.03)

§ Basic model - stratified by age of recruitment (1 year categories) and centre.

¶ Multivariable model - stratified by age of recruitment (1 year categories) and centre and adjusted for BMI, total energy intake, education, smoking status and intensity, and intakes of alcohol, fruits and vegetables, and meat and meat products.

Table 14. Assessment of prostate cancer incidence risk using the predictor 25-hydroxyvitamin D score (A) compared with a previous published analysis (97) (B) within the European Prospective Investigation into Cancer and Nutrition

Analyses were done in the full cohort containing 112,957 men.

(A)

Predicted 25(OH)D score	Q1	Q2	Q3	Q4	Q5		
	<42.8	42.8-45.6	45.7-48.1	48.2-50.9	≥51.0	*P*-trend	HR per 25 nmol/L
Men							
N cases	581	568	522	448	434		
Basic model §	1.00	1.10 (0.97-1.24)	1.07 (0.94-1.22)	0.99 (0.86-1.13)	1.09 (0.95-1.26)	0.53	1.09 (0.86-1.38)
Multivariable model ¶ minus BMI	1.00	1.08 (0.96-1.22)	1.05 (0.92-1.20)	0.97 (0.84-1.11)	1.07 (0.93-1.23)	0.70	1.05 (0.83-1.34)
Multivariable model ¶	1.00	1.09 (0.96-1.23)	1.06 (0.92-1.21)	0.98 (0.84-1.13)	1.08 (0.92-1.26)	0.70	1.07 (0.81-1.41)
Multivariable model ¶ plus physical activity	1.00	1.09 (0.95-1.25)	1.06 (0.90-1.25)	0.98 (0.81-1.19)	1.12 (0.90-1.39)	0.53	1.18 (0.80-1.72)

§ Basic model - stratified by age of recruitment (1 year categories) and centre.

¶ Multivariable model - stratified by age of recruitment (1 year categories) and centre and adjusted for BMI, education, smoking status and intensity, marital status, and intake of alcohol.

(B)

Travis *et al*., 2009		Standardised by month					
	Q1	Q2	Q3	Q4	Q5		
	≤40.4	>40.4-≤50.4	>50.4-≤59.1	>59.1-≤70.8	>70.8	*P*-trend	RR per 25 nmol/L
Matching variables only †	1.00	1.24 (0.88-1.74)	1.15 (0.82-1.62)	1.01 (0.71-1.45)	1.24 (0.87-1.79)	0.27	
Multivariable adjusted φ	1.00	1.27 (0.89-1.81)	1.23 (0.85-1.76)	1.06 (0.73-1.55)	1.28 (0.88-1.88)	0.19	1.05 (0.93-1.19)

† Matched by centre, age at recruitment, time of day of blood collection, and time between blood collection and last consumption of food and drink.

φ Multivariable model - adjusted for BMI, smoking, alcohol intake, education, marital status, and physical activity.

Sensitivity analyses for predictor 25-hydroxyvitamin D scores derivations

Various sensitivity analyses were carried out to assess the impact of certain decisions made during the derivation of the predictor 25(OH)D scores:

1) Due to smoking behaviour being an important risk factor/confounder in chronic disease/mortality risk models, as well as a significant correlate of circulating 25(OH)D levels, smoking status and intensity was excluded when the predictor scores were applied to the validation datasets to minimise the risk of statistical over-adjustment. In the sensitivity analysis, smoking status and intensity was excluded from the predictor scores and all validations were re-run.

2) Similarly, age is an important risk factor/confounder in chronic disease/mortality risk models, as well as an important determinant of circulating 25(OH)D levels. Age at blood collection was excluded when the predictor scores were applied to the validation datasets to minimise the risk of statistical over-adjustment. In the sensitivity analysis, age at blood collection was excluded from the predictor scores and all validations were re-run.

3) The predictor scores were derived using case and control samples from previous EPIC nested case-control studies. Cases were not excluded as blood samples were collected prospectively prior to the onset of disease. In sensitivity analyses, the predictor scores were derived using samples from control participants only to assess the impact of samples from cases being used.

Sensitivity analysis - correlations between actual and predicted 25-hydroxyvitamin D values

For the final men and women's predictor scores, the Pearson and Spearman correlation coefficients between actual and predicted 25(OH)D levels in the validation datasets ranged from 0.20 to 0.21. Similar correlations were observed when: (1) smoking status and intensity variable was excluded from the predictor scores at the derivation stage (Spearman: r =0.21 for men and r =0.20 for women); (2) age at blood collection was excluded from the predictor scores at the derivation stage (Spearman: r =0.20 for men and r =0.20 for women); and (3) when the predictor scores were derived using samples from control participants (Spearman: r =0.21 for men and r =0.19 for women).

Sensitivity analysis - actual circulating 25(OH)D measurements according to quintiles and deciles of predicted 25-hydroxyvitamin D levels

In the final predictor scores, actual circulating 25(OH)D measurements generally increased with increasing decile of predicted 25(OH)D levels in the validation datasets. The differences in mean actual 25(OH)D between the extreme predicted 25(OH)D deciles was 23.3 nmol/L for men and 15.2 nmol/L for women. Similar results were observed in the sensitivity analyses when: (1) the smoking status and intensity variable was excluded from the predictor scores at the derivation stage (*P*-trend <0.001; differences in extreme deciles 24.7 nmol/L for men and 15.1 nmol/L for women); and (2) age at blood collection was excluded from the predictor scores at the derivation stage (*P*-trend <0.001; differences in extreme deciles 24.7 nmol/L for men and 15.6 nmol/L for women). When samples from control participants were used to derive the predictor scores (3), once more, actual measurements increased with increasing decile of predicted 25(OH)D levels. However, compared to the final predictor scores, the difference between the extreme deciles was lower for men (19.8 nmol/L) but higher for women (20.1 nmol/L)

Sensitivity analysis - cross-classifications of participants into categories of predicted and actual 25-hydroxyvitamin D levels

In the final predictor scores, for men and women, 46.3% and 45.8% were respectively classified into equivalent or parallel quintiles of actual and predicted 25(OH)D in the validation datasets. Similar classifications were made when: (1) smoking status and intensity variable was excluded from the predictor scores at the derivation stage (45.6% men and 46.5% women); (2) age at blood collection was excluded from the predictor scores at the derivation stage (46.3% men and 46.0% women); and (3) when the predictor scores were derived using samples from control participants (44.5% men and 44.8% women).

Sensitivity analysis - assessment of colorectal cancer incidence risk using predicted 25-hydroxyvitamin D score

Colorectal cancer incidence risk was assessed using the predictor scores in the nested case-control datasets (with the circulating samples sourced from the colorectal cancer nested case-control study excluded). In the final predictor score sexes combined multivariable model, a 20% lower colorectal cancer risk (OR 0.80, 95% CI: 0.68-0.93) was observed per increment of 5.9 nmol/L (100 IU/L). In sensitivity analyses, similar inverse associations were observed when: (1) smoking status and intensity variable was excluded from the predictor scores at the derivation stage (HR per 5.9 nmol/L increment in predicted

25(OH)D, 0.80, 95% CI: 0.68-0.93); (2) age at blood collection was excluded from the predictor scores at the derivation stage (HR per 5.9 nmol/L increment in predicted 25(OH)D, 0.77, 95% CI: 0.66-0.91); and (3) when the predictor scores were derived using samples from control participants (HR per 5.9 nmol/L increment in predicted 25(OH)D, 0.80, 95% CI: 0.68-0.93) (Table 15). Similar associations were observed for all versions of the predictor scores when not adjusted for BMI in the multivariable models, and when additionally adjusted for physical activity index.

Table 15. Assessment of colorectal cancer incidence risk in the colorectal nested case-control dataset using the: 25-hydroxyvitamin D final predictor scores, (1) final predictor scores minus smoking, (2) final predictor scores minus age at blood collection, and (3) final predictor scores using control samples only

	RR per 100 IU/L (5.9 nmol/L)			
Predicted 25(OH)D score	Final predictor scores	Smoking excluded from predictor scores [1]	Age at blood collection excluded from predictor scores [2]	Control samples only predictor scores [3]
Multivariable model ϕ minus BMI	0.79 (0.70-0.89)	0.79 (0.70-0.89)	0.78 (0.69-0.88)	0.79 (0.70-0.89)
Multivariable model ϕ	0.80 (0.68-0.93)	0.80 (0.68-0.93)	0.77 (0.66-0.91)	0.80 (0.68-0.93)
Multivariable model ϕ plus physical activity adjustment	0.79 (0.66-0.95)	0.79 (0.65-0.95)	0.76 (0.62-0.93)	0.79 (0.66-0.95)

† Matched by centre, sex, age at recruitment, and time of day of blood collection.

ϕ Multivariable model - adjusted for BMI, smoking, alcohol intake, education, physical activity, total energy intake, and intakes of total fruits and vegetables, meats/meat products, and alcohol.

1 Excluding smoking status and intensity from predictor score derivation.

2 Excluding age at blood collection from predictor score derivation.

3 Predictor scores derived using 25(OH)D samples from control participants only (i.e. excluding cases from lymphoma and prostate cancer nested case-control studies).

3.3 Application and validation of the Health Professionals Follow-Up Study derived 25-hydroxyvitamin D score in the European Prospective Investigation into Cancer and Nutrition

3.3.1 Validation of the Health Professionals Follow-Up Study derived 25-hydroxyvitamin D score in the European Prospective Investigation into Cancer and Nutrition

The U.S. based HPFS predictor score (52) was applied to the EPIC validation dataset of 823 men who lived at similar latitudes. Application of the predictor score created a predicted 25(OH)D value for each participant.

Correlations between actual and predicted 25-hydroxyvitamin D values
The mean (SD) of the HPFS predicted 25(OH)D score was 69.4 (6.7) nmol/L; compared to mean (SD) of actual circulating 25(OH)D measurements (adjusted for course source study and batch of laboratory analysis, month of blood collection, and age of blood collection) of 57.2 (22.7) nmol/L. The correlations between actual circulating 25(OH)D measurements and the HPFS score were lower than for the EPIC derived score; with identical Pearson and Spearman correlation coefficients (0.05, 95% CI: 0.02-0.12) yielded between the HPFS predicted 25(OH)D values and actual circulating 25(OH)D measurements.

Actual circulating 25(OH)D measurements according to quintiles and deciles of Health Professional Follow-Up Study predicted 25-hydroxyvitamin D levels
Mean actual circulating 25(OH)D levels were changeable across increasing quintiles and deciles of HPFS predicted 25(OH)D levels (Figure 10). Unlike the EPIC derived predictor scores, linear increases in actual 25(OH)D across quintiles (*P*-trend 0.17) and deciles (*P*-trend 0.18) of HPFS predicted 25(OH)D were not observed. Furthermore, the differences across the extreme HPFS predicted quantiles of mean actual 25(OH)D were narrower than from the EPIC derived score (5.0 nmol/L for quintiles vs. 18.3 nmol/L using the EPIC derived score; and 5.8 nmol/L for deciles vs. 23.3 nmol/L using the EPIC derived score).

Cross-classifications of participants into categories of Health Professional Follow-Up Study predicted and actual 25-hydroxyvitamin D levels

Figure 11 shows the proportion of participants classified by quintiles of actual and HPFS predicted 25(OH)D. Of men classified in the first actual circulating 25(OH)D quintile: 25.9% were classified into predicted 25(OH)D quintile 1; 44.6% were classified into predicted 25(OH)D quintile 1 or 2; and 18.1% were classified into predicted 25(OH)D quintile 5. Of those men classified into the highest actual circulating 25(OH)D quintile: 21.5% were classified into the corresponding predicted 25(OH)D quintile; 36.8% were classified either in quintiles 4 and 5 of predicted 25(OH)D; 27.0% were classified into predicted quintile 3; and 17.2% were classified into the lowest predicted 25(OH)D quintile. Overall, 42.1% of men were classified into equivalent or parallel quintiles of actual and predicted 25(OH)D.

Assessment of colorectal cancer incidence risk using the Health Professionals Follow-Up Study predicted 25-hydroxyvitamin D score

The next stage in the validation process of the predicted 25(OH)D scores was to assess the risks of colorectal cancer and then compare these results against previously published studies with EPIC. Unlike the inverse associations observed in the Jenab *et al.,* (67) and when using the EPIC derived predictor score, non-significant associations were observed when the HPFS predictor score was applied to the full EPIC cohort dataset to assess colorectal cancer incidence risk (multivariable model: HR per 5.9 nmol/L increment, 1.08, 95% CI: 0.88-1.33). Similarly, in the categorical models, when the highest and lowest quintiles were compared, a non-significant 23% increased risk (HR 1.23, 95% CI: 0.76-1.99) was observed, without a linear trend between quintiles (*P*-trend 0.49). Analysis of the HPFS predictor score within the colorectal cancer nested case-control dataset was not possible as, after exclusions (including circulating samples sourced from the colorectal cancer nested case-control study), just 4 cases remained in the dataset of 558 men.

Figure 10. Mean actual circulating 25-hydroxyvitamin D level by: (A) quintile; (B) decile of predicted 25-hydroxyvitamin D score derived in the Health Professionals Follow-Up Study but applied in the European Prospective Investigation into Cancer and Nutrition

Analyses were done in the validation dataset of 823 men.

(A)
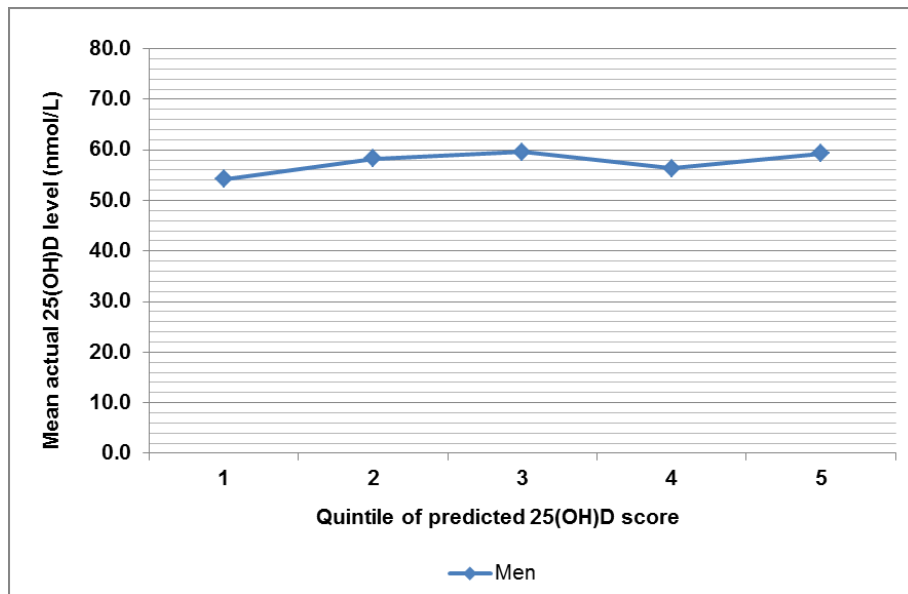
(B)
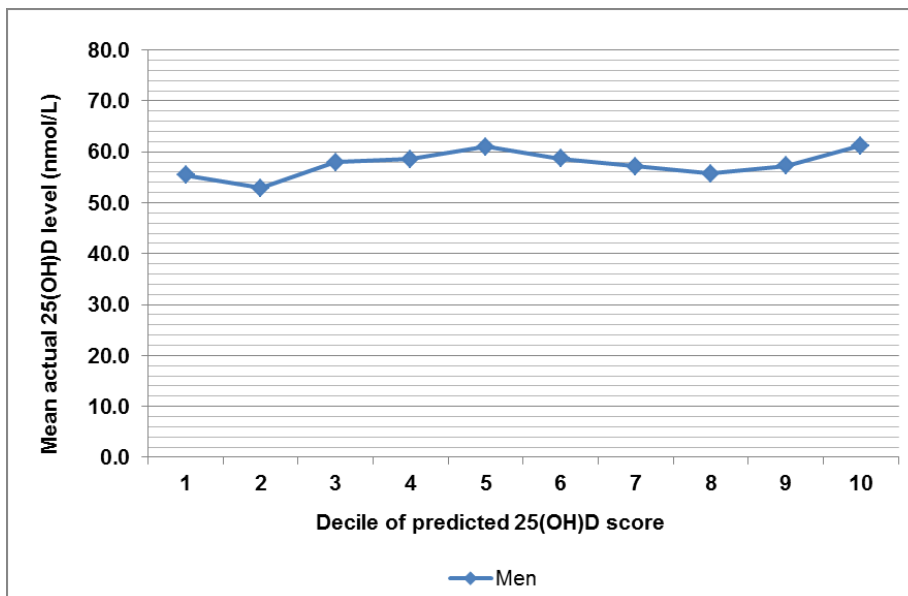
Figure 11. The percentage of participants who were classified into predicted circulating 25-hyrdroxyvitamin D by their actual circulating 25(OH)D quintiles using the Health Professionals Follow-Up Study predictor score

Analyses were done in the validation dataset of 823 men.
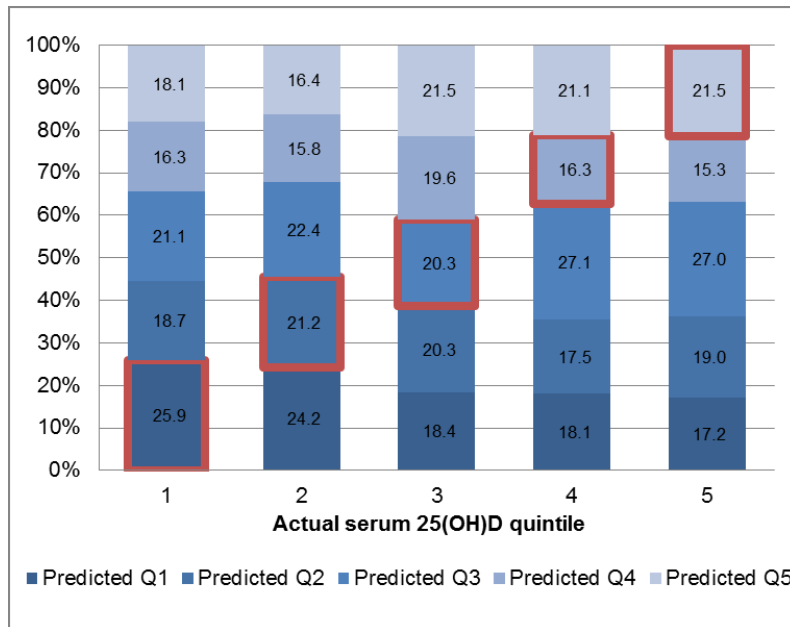
3.4 Application of the predicted 25-hydroxyvitamin D score within the European Prospective Investigation into Cancer and Nutrition to assess cancer incidence risk

3.4.1 Characteristics of cohort participants

The cohort characteristics by country and sex are shown in Table 16. The highest predicted circulating 25(OH)D levels were observed in the Netherlands for men and Denmark for women. The lowest predicted 25(OH)D levels for men and women were observed amongst participants from Germany and Greece respectively. The crude overall cancer incidence rates for men and women were 80 and 72 cases per 10,000 person-years respectively. After adjustment for age, using the European standard population (175), incidence rates for men and women were 67 and 58 cases per 10,000 person-years respectively.

Table 16. Cohort characteristics by sex and country of participants included in the predicted 25(OH)D-cancer incidence analyses

Country	N of participants		Total person-years		N overall cancer cases		Mean (SD) predicted 25(OH)D level (nmol/L)	
	Men	Women	Men	Women	Men	Women	Men	Women
Denmark	25,504	27,945	278,258	309,429	3,086	3,240	48.0 (4.0)	65.0 (6.8)
France	-	19,087	-	204,184	-	1,976	-	54.4 (4.8)
Germany	20,457	27,017	203,323	268,491	1,881	1,915	42.7 (4.2)	49.1 (5.8)
Greece	10,149	14,468	93,279	141,453	473	507	44.8 (3.7)	44.1 (5.2)
Italy	13,472	29,977	153,720	336,257	1,003	2,286	48.5 (3.7)	49.3 (5.1)
Spain	14,737	24,532	179,334	296,363	1,089	1,271	46.7 (4.5)	56.3 (6.9)
The Netherlands	7,342	23,668	84,805	273,899	354	2,129	50.7 (4.0)	56.2 (5.0)
United Kingdom	21,296	50,773	237,872	567,815	2,009	3,861	48.8 (4.3)	55.7 (5.3)
All EPIC	112,957	217,467	1,230,590	2,397,890	9,895	17,185	47.0 (4.8)	54.4 (7.9)

3.4.2 Baseline characteristics by predicted 25-hydroxyvitamin D quintiles

The characteristics of the participants included in the analyses of cancer incidence by quintiles of predicted 25(OH)D are shown in Table 17. Compared to participants in the lowest quintile, those participants in the highest quintile of predicted 25(OH)D were more likely to be younger, physically active, and have a lower BMI. Men in the highest predicted 25(OH)D quintile had lower intakes of red and processed meat and higher intakes of total energy and fruits and vegetables than those in the lowest quintile. Conversely, women in the

highest predicted 25(OH)D quintile had higher red and processed meat intakes and lower intakes of total energy and fruits and vegetables than those in the lowest quintile. Women in the highest predicted 25(OH)D quintile were more likely to be current smokers than those in the lowest quintile. Current smokers were relatively evenly distributed in men across the predicted 25(OH)D quintiles. Women in the lowest predicted 25(OH)D quintile were less likely to have used the contraceptive pill than those in the higher quintiles.

Table 17. Baseline characteristics of study participants included in the cancer incidence analyses by categories (quintiles) of predicted 25-hydroxyvitamin D score

Characteristic	Quintile of predicted 25(OH)D (nmol/L)				
	Q1	Q2	Q3	Q4	Q5
Men					
Predicted 25(OH)D range (nmol/L)	<42.8	42.8-45.7	45.8-48.1	48.2-50.9	≥51.0
N	22,592	22,591	22,592	22,591	22,591
N overall cancer cases	2397	2118	1993	1813	1574
Age at recruitment (years)	55.2 (8.8)	53.7 (9.2)	52.4 (9.3)	51.1 (9.6)	48.7 (10.8)
Body mass index (kg/m2)	29.3 (3.9)	27.4 (3.4)	26.6 (3.1)	25.6 (2.9)	24.3 (2.8)
Education ‡ Longer education including University (%)	29	27.6	27.3	28.5	27.6
Smoking status and intensity ‡ Current (%)	30.1	31.9	31.4	31.2	30.4
Physical activity ‡ Active (%)	2.9	9.7	17.5	32.9	67.9
Total energy intake (kcal/day)	2334 (671)	2389 (662)	2427 (643)	2454 (644)	2511 (660)
Red and processed meat consumption (g/day)	108.1 (65.4)	101.0 (61.4)	98.3 (60.8)	97.0 (62.1)	94.6 (63.5)
Fruit & vegetable consumption (g/day)	392.5 (299.0)	436.9 (307.2)	448.8 (293.2)	439.2 (274.7)	432.0 (265.3)
Vitamin D intake (µg/day)	3.2 (1.8)	3.7 (2.1)	3.9 (2.2)	4.3 (2.4)	5.5 (3.9)
Calcium intake (mg/day)	943.9 (405.0)	1003 (414.6)	1044 (413.4)	1076 (421.3)	1124 (438.7)
Alcohol intake (g/day)	24.3 (27.8)	23.3 (24.9)	23.1 (23.7)	22.3 (22.3)	19.7 (20.6)
Women					
Predicted 25(OH)D range (nmol/L)	<48.0	48.0-52.4	52.5-56.1	56.2-60.3	≥60.4
N	43,494	43,493	43,494	43,493	43,493
N overall cancer cases	3260	3495	3331	3273	3826
Age at recruitment (years)	54.2 (10.1)	50.9 (10.3)	49.2 (10.6)	47.8 (11.3)	51.3 (9.8)
Body mass index (kg/m2)	29.9 (5.1)	26.0 (4.1)	24.5 (3.7)	23.6 (3.5)	23.4 (3.2)
Education ‡ Longer education including University (%)	12.5	21.4	26.8	29.0	19.4
Smoking status and intensity ‡ Current (%)	17.1	18.4	17.4	18.0	25.9
Physical activity ‡ Active (%)	4.2	9.6	14.9	22.9	34.2
Total energy intake (kcal/day)	1950 (575)	1962 (549)	1957 (535)	1935 (510)	1921 (482)
Red and processed meat consumption (g/day)	65.9 (42.9)	64.8 (43.2)	61.0 (44.9)	57.3 (44.7)	67.0 (40.4)
Fruit & vegetable consumption (g/day)	530.3 (306.4)	483.2 (283.9)	486.0 (264.3)	487.4 (264.0)	451.0 (246.2)
Vitamin D intake (µg/day)	2.5 (1.5)	2.8 (1.6)	3.0 (1.7)	3.3 (1.8)	4.5 (2.7)
Calcium intake (mg/day)	947.2 (377)	993.4 (395)	1019 (394)	1032 (388)	1041 (409)
Alcohol intake (g/day)	6.4 (11.3)	8.1 (11.9)	8.6 (11.8)	8.6 (11.5)	10.5 (12.6)
Ever use of contraceptive pill ‡ Yes (%)	37.7	54.8	63.7	68.5	64.1
Menopausal status ‡ Postmenopausal (%)	54.7	42.7	37.2	36.3	52.1

Mean and standard deviation (in parenthesis) for continuous variables, or percentages for categorical variables (‡).

3.4.3 Predicted 25-hydroxyvitamin D levels and overall cancer incidence risk

After a mean (SD) follow-up of 11.0 (2.6) years, 27,080 incident cases of any cancer accrued. In the basic models (adjusted for age, sex, and centre), a linear (*P*-trend <0.001) inverse association was observed between predicted 25(OH)D level and overall cancer incidence risk in the men and women's combined analysis (Table 18). This association attenuated after multivariable adjustments (Q5 vs. Q1, HR 0.89, 95% CI 0.84-0.94), but a linear relationship remained (*P*-trend <0.001). In the continuous model, a 13% lower overall cancer incidence risk was observed per increment in predicted 25 nmol/L of 25(OH)D. After adjustment for physical activity index, the associations attenuated further, and statistical significance was lost in the continuous, but not the categorical models. Similar significant inverse associations were observed when participants whose follow-up time was less than 5 years were excluded from the analyses. The statistically significant association observed amongst men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.81, 95% CI: 0.70-0.93) was slightly stronger than that observed amongst women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.88, 95% CI: 0.81-0.96); however, this difference was not statistically significant (*P*-interaction 0.12).

Obesity is both a risk factor for cancer and correlated with vitamin D status. To examine whether the associations were driven largely by obesity, models with and without predicted circulating 25(OH)D variables were run. Similar associations for BMI were observed with (BMI 35+ vs. 22-24.9 kg/m^2, HR 1.13, 95% CI: 1.06-1.22) and without (BMI 35+ vs. 22-24. 9 kg/m^2, HR 1.18, 95% CI: 1.11-1.27) predicted 25(OH)D in the multivariable model. This indicates that the predicted vitamin D-overall cancer incidence association is largely independent of obesity.

No statistically significant interactions were observed for predicted 25(OH)D levels and overall cancer incidence across strata of lifestyle, demographic, anthropometric, and dietary variables (Table 19). The inverse associations were constant across strata of age, follow-up time, menopausal status, ever use of contraceptive pills and dietary intakes of red and processed meats, fruits and vegetables, fish, calcium, fibre, and retinol. For alcohol, a stronger and significant reduced risk was observed amongst non-consumers, although this difference to the other consumption categories was non-significant (*P*-interaction 0.12). No association was observed amongst never smokers which may indicate possible residual

confounding; because of this the associations between predicted 25(OH)D levels and smoking and non-smoking cancers were investigated.

The mutually adjusted HRs for total overall cancer incidence associated with the individual components of the applied predictor scores are shown in Table 20. After multivariable adjustments, waist circumference, country, and ever use of menopausal hormone therapy were significantly associated with overall cancer risk. In contrast, dietary vitamin D intake and physical activity were not significantly associated with overall cancer incidence risk

The associations between predicted 25(OH)D and overall cancer incidence by country are shown in Figure 12. The inverse association for overall cancer incidence was relatively consistent across countries (*P*-heterogeneity 0.08), with only Germany yielding a risk estimate >1.

Table 18. Risk (hazard ratios) of overall cancer incidence associated with predicted 25-hydroxyvitamin D level

	Quintile of predicted 25(OH)D (nmol/L)							<5 years excluded
	1	2	3	4	5		Per 25 nmol/L increase	Per 25 nmol/L increase
Men	<42.8	42.8-45.6	45.7-48.1	48.2-50.9	≥51.0			
Women	<48.0	48.0-52.4	52.5-56.1	56.2-60.3	≥60.4	*P*-trend		
Overall cancer incidence								
Both sexes								
N cases	5,657	5,613	5,324	5,086	5,400			
Person-years	686,034	716,725	734,411	743,941	747,370			
Basic model - HR (95% CI) †	1.00	0.94 (0.90-0.98)	0.89 (0.86-0.93)	0.87 (0.84-0.91)	0.84 (0.80-0.88)	<0.001	0.80 (0.76-0.85)	0.77 (0.72-0.83)
Multivariable model - HR (95% CI) ‡	1.00	0.97 (0.93-1.01)	0.93 (0.89-0.97)	0.91 (0.87-0.96)	0.89 (0.84-0.94)	<0.001	0.87 (0.80-0.92)	0.84 (0.77-0.92)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.98 (0.94-0.94)	0.95 (0.90-0.99)	0.94 (0.89-0.99)	0.92 (0.86-0.99)	0.012	0.92 (0.84-1.01)	0.85 (0.76-0.95)
Men								
N cases	2,397	2,118	1,993	1,813	1,574			
Person-years	229,777	241,938	249,081	252,868	256,927			
Basic model - HR (95% CI) *	1.00	0.91 (0.86-0.97)	0.88 (0.82-0.94)	0.85 (0.79-0.91)	0.82 (0.77-0.89)	<0.001	0.69 (0.61-0.78)	
Multivariable model - HR (95% CI) †	1.00	0.95 (0.89-1.01)	0.94 (0.88-1.01)	0.91 (0.85-0.98)	0.90 (0.83-0.97)	0.006	0.81 (0.70-0.93)	
Multivariable model + physical activity adj - HR (95% CI) †	1.00	0.96 (0.90-1.03)	0.96 (0.88-1.04)	0.94 (0.85-1.03)	0.93 (0.83-1.04)	0.21	0.86 (0.70-1.06)	
Women								
N cases	3,260	3,495	3,331	3,273	3,826			
Person-years	456,257	474,787	485,330	491,073	490,443			
Basic model - HR (95% CI) *	1.00	0.96 (0.91-1.01)	0.90 (0.86-0.95)	0.89 (0.84-0.94)	0.86 (0.80-0.91)	<0.001	0.84 (0.79-0.90)	
Multivariable model - HR (95% CI) †	1.00	0.98 (0.93-1.03)	0.92 (0.87-0.98)	0.91 (0.85-0.97)	0.89 (0.82-0.96)	0.001	0.88 (0.81-0.96)	
Multivariable model + physical activity adj - HR (95% CI) †	1.00	0.99 (0.94-1.05)	0.95 (0.89-1.01)	0.95 (0.88-1.02)	0.93 (0.85-1.01)	0.06	0.94 (0.85-1.04)	

† Basic model - Cox regression stratified by age (1-year categories), sex, and centre.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

φ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Table 19. Risk (hazard ratios) of overall cancer incidence associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level across strata of lifestyle, demographic, anthropometric, and dietary variables

	Overall cancer incidence	
	Both sexes	P interaction
Stratification variable	Predicted 25(OH)D	
	HR (95% CI), per 25 nmol/L ‡	
Overall	0.87 (0.80-0.92)	
Sex		0.12
Men	0.81 (0.70-0.93)	
Women	0.88 (0.81-0.96)	
Age at recruitment		0.15
<55 years old	0.86 (0.77-0.97)	
55-<65 years old	0.89 (0.80-0.99)	
≥65 years old	0.76 (0.60-0.96)	
Follow-up		0.40
<5 years	0.93 (0.81-1.05)	
5-10 years	0.90 (0.81-1.00)	
≥10 years	0.81 (0.69-0.95)	
Smoking status		0.15
Never smoked	0.98 (0.87-1.09)	
Former smoker	0.79 (0.69-0.91)	
Current smoker	0.70 (0.61-0.81)	
Body mass index		0.12
<24.9 kg/m2	0.94 (0.84-1.05)	
<25.0-29.9 kg/m2	0.84 (0.75-0.95)	
≥30 kg/m2	0.69 (0.58-0.82)	
Physical activity		0.06
Inactive	0.75 (0.62-0.92)	
Moderately inactive	1.01 (0.86-1.18)	
Moderately active	1.06 (0.88-1.27)	
Active	0.88 (0.72-1.08)	
Alcohol consumption		0.12
Non-consumers	0.66 (0.54-0.82)	
<15 g/day	0.90 (0.82-0.99)	
15-29.9 g/day	0.97 (0.79-1.19)	
≥30 g/day	0.86 (0.71-1.04)	
Red and processed meat consumption *		0.38
Below median	0.92 (0.83-1.03)	
Above median	0.82 (0.74-0.91)	
Fruit and vegetable consumption *		0.31
Below median	0.85 (0.77-0.94)	
Above median	0.88 (0.79-0.99)	
Fish consumption *		0.98
Below median	0.73 (0.65-0.84)	
Above median	0.91 (0.83-0.99)	
Calcium intake *		0.99
Below median	0.85 (0.76-0.94)	
Above median	0.90 (0.81-0.99)	
Fibre intake *		0.92
Below median	0.84 (0.75-0.93)	
Above median	0.90 (0.81-1.01)	
Retinol intake *		0.69
Below median	0.87 (0.78-0.98)	
Above median	0.86 (0.78-0.95)	
Menopausal status		0.28
Premenopausal	0.85 (0.70-1.04)	
Postmenopausal	0.90 (0.81-1.01)	
Perimenopausal	0.85 (0.66-1.09)	
Surgical postmenopausal	0.74 (0.48-1.14)	
Ever use of contraceptive pill		0.68
No	0.88 (0.78-1.01)	
Yes	0.87 (0.77-0.99)	

* Median intakes: red and processed meat=93.3 g/day in men and 59.0 g/day in women; fruit and vegetable=356 g/day in men and 438 g/day in women; fish consumption=27.1 g/day in men and 21.9 g/day in women; calcium=972 mg/day in men and 952 mg/day in women; fibre=23.9 g/day in men and 21.7 g/day in women; retinol=691 µg/day in men and 509 µg/day in women.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school

completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Table 20. Risk (hazard ratios) of overall cancer incidence associated with the correlates/determinants of circulating 25(OH)D used in the predictor scores (mutually adjusted)

	Overall cancer	
	Both sexes	
Predictor score	HR (95% CI)	P-trend
Dietary vitamin D (Per 2.5 µg/day)	1.00 (0.99-1.02)	
Physical activity		0.94
Inactive	1.00	
Moderately inactive	0.99 (0.96-1.02)	
Moderately active	1.01 (0.97-1.05)	
Active	0.99 (0.95-1.03)	
Waist circumference (Per 1 cm)	1.01 (1.00-1.02)	
Country		0.01
Denmark	1.00	
UK	0.97 (0.93-1.02)	
The Netherlands	0.91 (0.87-0.96)	
France	1.22 (1.14-1.30)	
Germany	1.13 (1.08-1.18)	
Greece	0.49 (0.45-0.53)	
Italy	0.94 (0.90-0.99)	
Spain	0.77 (0.72-0.81)	
Ever use of menopausal hormone therapy		<0.001
No	1.00	
Yes	1.18 (1.13-1.22)	

Multivariable model - Cox regression with predictor score determinants mutually adjusted for each other, plus additional adjustment for body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Figure 12. Risk (hazard ratios) of overall cancer incidence, by country, associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level

	Number of cases		HR (95% CI) per predicted 25 nmol/L increase
Denmark	6326		0.90 (0.78, 1.03)
France	1976		0.94 (0.69, 1.29)
Germany	3796		1.05 (0.87, 1.27)
Greece	980		0.60 (0.36, 1.02)
Italy	3289		0.67 (0.52, 0.86)
Spain	2360		0.92 (0.74, 1.14)
The Netherlands	2483		0.66 (0.49, 0.88)
UK - General population	5218		0.85 (0.72, 0.99)
UK - Health conscious	652		0.86 (0.55, 1.37)
ALL EPIC	27080		0.87 (0.80, 0.92)

Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

3.4.4 Predicted 25-hydroxyvitamin D levels and smoking related and non-smoking related cancer incidences

Smoking related cancers

The associations between predicted 25(OH)D and incidence of smoking related cancers are shown in Table 21. In the sexes combined multivariable model, a 24% reduced smoking related cancer risk (95% CI: 0.69-0.83) was observed when the highest and lowest quintiles of predicted circulating 25(OH)D were compared (*P*-trend <0.001). In the equivalent continuous model, a 35% lower smoking related cancer risk (95% CI: 0.56-0.74) was observed per 25 nmol/L higher predicted 25(OH)D. The association was similar when the first 5 years of follow-up were excluded. When analysed by sex, similar strength inverse associations were observed (*P*-interaction 0.35) amongst men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.67, 95% CI: 0.55-0.82) and women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.61, 95% CI: 0.51-0.73) (Table 21).

In sub-group analyses by strata of lifestyle, demographic, anthropometric, and dietary variables, a significant interaction was observed only for smoking status (*P*-interaction 0.01) (Table 22). This may indicate residual confounding by smoking impacted upon the results. However, significant inverse associations were observed across all categories of smoking status, including amongst never smokers, where a 29% lower (95% CI: 0.55-0.92) risk was observed per 25 nmol/L increment in predicted 25(OH)D. Inverse associations were observed across all countries (*P*-heterogeneity 0.14); with significant associations observed in Denmark, the Netherlands, and the UK (general population) cohorts (Figure 13A).

Non-smoking related cancers

No associations were observed for non-smoking related cancers in the sexes combined categorical (Q5 vs. Q1, HR 0.98, 95% CI 0.91-1.05; *P*-trend 0.26) and continuous (HR per 25 nmol/L increment in predicted 25(OH)D, 0.98, 95% CI: 0.89-1.07) models (Table 21). Similar non-significant risk estimates were observed when men and women were analysed separately (*P*-interaction 0.06).

Non-significant estimates were observed across all countries (*P*-heterogeneity 0.38) (Figure 13B). The heterogeneous relationships between predicted 25(OH)D and smoking and non-smoking related cancers was significant (*P*-heterogeneity <0.001).

Table 21. Risk (hazard ratios) of smoking related and non-smoking related cancer incidences associated with predicted 25-hydroxyvitamin D level

	Quintile of predicted 25(OH)D (nmol/L)							<5 years excluded
	1	2	3	4	5		Per 25 nmol/L	Per 25 nmol/L
Men	<42.8	42.8-45.6	45.7-48.1	48.2-50.9	≥51.0		increase	increase
Women	<48.0	48.0-52.4	52.5-56.1	56.2-60.3	≥60.4	*P*-trend		
Smoking related cancers								
Both sexes								
N cases	2,125	1,808	1,584	1,513	1,626			
Basic model - HR (95% CI) †	1.00	0.86 (0.80-0.92)	0.78 (0.72-0.84)	0.77 (0.71-0.83)	0.69 (0.63-0.75)	<0.001	0.59 (0.52-0.66)	0.54 (0.47-0.62)
Multivariable model - HR (95% CI) ‡	1.00	0.90 (0.84-0.97)	0.84 (0.78-0.91)	0.84 (0.77-0.91)	0.76 (0.69-0.83)	<0.001	0.65 (0.56-0.74)	0.63 (0.54-0.74)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.93 (0.86-0.99)	0.88 (0.81-0.96)	0.88 (0.80-0.97)	0.80 (0.71-0.90)	<0.001	0.69 (0.58-0.82)	0.66 (0.54-0.81)
Men								
N cases	1283	1058	945	887	706			
Multivariable model - HR (95% CI) ‡	1.00	0.92 (0.84-1.00)	0.87 (0.79-0.96)	0.88 (0.80-0.98)	0.81 (0.72-0.91)	<0.001	0.67 (0.55-0.82)	
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.94 (0.85-1.03)	0.90 (0.80-1.02)	0.91 (0.80-1.04)	0.83 (0.71-0.97)	0.04	0.67 (0.50-0.92)	
Women								
N cases	842	750	639	626	920			
Multivariable model - HR (95% CI) ‡	1.00	0.88 (0.78-0.98)	0.79 (0.69-0.89)	0.75 (0.66-0.87)	0.66 (0.56-0.78)	<0.001	0.61 (0.51-0.73)	
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.90 (0.80-1.01)	0.82 (0.72-0.94)	0.80 (0.68-0.93)	0.71 (0.59-0.86)	<0.001	0.66 (0.53-0.82)	
Non-smoking related cancers								
Both sexes								
N cases	3,532	3,805	3,740	3,573	3,774			
Basic model - HR (95% CI) †	1.00	0.99 (0.94-1.04)	0.96 (0.91-1.01)	0.93 (0.88-0.99)	0.94 (0.88-0.99)	0.01	0.92 (0.85-0.98)	0.91 (0.83-0.99)
Multivariable model - HR (95% CI) ‡	1.00	1.01 (0.96-1.06)	0.98 (0.93-1.04)	0.96 (0.91-1.02)	0.98 (0.91-1.05)	0.26	0.98 (0.89-1.07)	0.96 (0.86-1.07)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.02 (0.96-1.07)	1.00 (0.94-1.06)	0.98 (0.92-1.05)	1.00 (0.92-1.09)	0.82	1.03 (0.93-1.14)	0.95 (0.83-1.09)
Men								
N cases	1114	1060	1048	926	868			
Multivariable model - HR (95% CI) ‡	1.00	0.98 (0.90-1.08)	1.01 (0.91-1.11)	0.94 (0.85-1.05)	0.98 (0.88-1.10)	0.61	0.96 (0.79-1.17)	
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.98 (0.89-1.08)	1.01 (0.90-1.13)	0.95 (0.83-1.09)	1.02 (0.88-1.19)	0.88	1.05 (0.80-1.39)	
Women								
N cases	2418	2745	2692	2647	2906			
Multivariable model - HR (95% CI) ‡	1.00	1.01 (0.95-1.08)	0.97 (0.90-1.04)	0.96 (0.89-1.04)	0.97 (0.89-1.05)	0.27	0.98 (0.89-1.09)	
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.03 (0.96-1.10)	0.99 (0.92-1.07)	0.99 (0.91-1.08)	1.00 (0.91-1.10)	0.81	1.05 (0.93-1.17)	

† Basic model - Cox regression stratified by age (1-year categories), sex, and centre.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever

use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

ϕ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Smoking related cancers were lung, bladder, upper aero-digestive, kidney, stomach, pancreatic, liver, colon, and rectum. Non-smoking related cancers were skin (non-melanoma), breast, endometrial, ovarian, prostate, brain, spinal cord/central nervous system, and thyroid.

Table 22. Risk (hazard ratios) of smoking related cancer incidence associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level across strata of lifestyle, demographic, anthropometric, and dietary variables

	Smoking related cancer incidence	
	Both sexes	P interaction
Stratification variable	Predicted 25(OH)D	
	HR (95% CI), per 25 nmol/L ‡	
Overall	0.65 (0.56-0.74)	
Sex		0.62
Men	0.67 (0.55-0.82)	
Women	0.61 (0.51-0.73)	
Age at recruitment		0.83
<55 years old	0.68 (0.54-0.86)	
55-<65 years old	0.64 (0.54-0.77)	
≥65 years old	0.61 (0.41-0.90)	
Follow-up		0.80
<5 years	0.72 (0.56-0.92)	
5-10 years	0.65 (0.53-0.79)	
≥10 years	0.70 (0.54-0.92)	
Smoking status		0.01
Never smoked	0.71 (0.55-0.92)	
Former smoker	0.58 (0.45-0.75)	
Current smoker	0.53 (0.43-0.65)	
Body mass index		0.23
<24.9 kg/m2	0.76 (0.61-0.94)	
<25.0-29.9 kg/m2	0.60 (0.49-0.75)	
≥30 kg/m2	0.54 (0.40-0.74)	
Physical activity		0.77
Inactive	0.59 (0.42-0.85)	
Moderately inactive	0.84 (0.62-1.15)	
Moderately active	0.65 (0.44-0.96)	
Active	0.70 (0.48-1.03)	
Alcohol consumption		0.63
Non-consumers	0.55 (0.37-0.81)	
<15 g/day	0.62 (0.51-0.75)	
15-29.9 g/day	0.93 (0.65-1.32)	
≥30 g/day	0.62 (0.46-0.84)	
Red and processed meat consumption *		0.32
Below median	0.68 (0.55-0.84)	
Above median	0.64 (0.54-0.77)	
Fruit and vegetable consumption *		0.13
Below median	0.66 (0.55-0.78)	
Above median	0.64 (0.52-0.79)	
Fish consumption *		0.97
Below median	0.49 (0.39-0.62)	
Above median	0.70 (0.59-0.83)	
Calcium intake *		0.11
Below median	0.69 (0.57-0.83)	
Above median	0.62 (0.51-0.75)	
Fibre intake *		0.07
Below median	0.68 (0.57-0.82)	
Above median	0.62 (0.51-0.76)	
Retinol intake *		0.86
Below median	0.67 (0.54-0.83)	
Above median	0.63 (0.53-0.75)	
Menopausal status		0.13
Premenopausal	0.50 (0.28-0.92)	
Postmenopausal	0.60 (0.48-0.74)	
Perimenopausal	0.53 (0.29-0.95)	
Surgical postmenopausal	0.98 (0.47-2.08)	
Ever use of contraceptive pill		0.48
No	0.62 (0.48-0.80)	
Yes	0.59 (0.45-0.78)	

* Median intakes: red and processed meat=93.3 g/day in men and 59.0 g/day in women; fruit and vegetable=356 g/day in men and 438 g/day in women; fish consumption=27.1 g/day in men and 21.9 g/day in women; calcium=972 mg/day in men and 952 mg/day in women; fibre=23.9 g/day

in men and 21.7 g/day in women; retinol=691 μg/day in men and 509 μg/day in women.

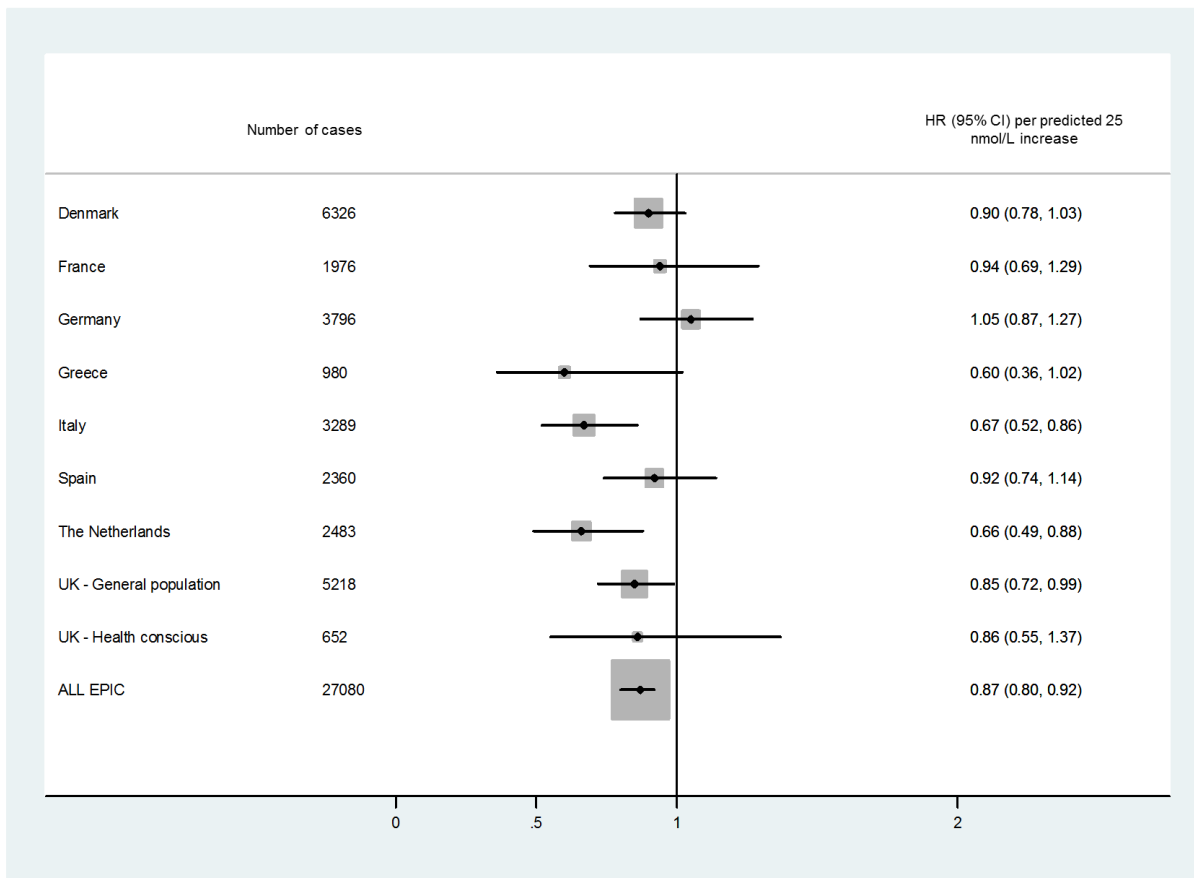‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Smoking related cancers were lung, bladder, upper aero-digestive, kidney, stomach, pancreatic, liver, colon, and rectum. Non-smoking related cancers (all other cancers).

Figure 13. Risk (hazard ratios) of smoking related (A) and non-smoking related (B) cancer incidences, by country, associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level

(A)

	Number of cases		HR (95% CI) per predicted 25 nmol/L increase
Denmark	2449		0.69 (0.54, 0.87)
France	169		0.36 (0.13, 1.02)
Germany	1339		0.75 (0.53, 1.07)
Greece	447		0.55 (0.24, 1.25)
Italy	1058		0.49 (0.31, 0.78)
Spain	861		0.88 (0.60, 1.30)
The Netherlands	749		0.35 (0.21, 0.59)
UK - General population	1450		0.72 (0.52, 0.99)
UK - Health conscious	134		0.50 (0.18, 1.41)
ALL EPIC	8656		0.65 (0.56, 0.74)

(B)

	Number of cases		HR (95% CI) per predicted 25 nmol/L increase
Denmark	3877		1.03 (0.86, 1.22)
France	1807		1.04 (0.75, 1.44)
Germany	2457		1.19 (0.97, 1.45)
Greece	533		0.66 (0.33, 1.33)
Italy	2231		0.78 (0.57, 1.06)
Spain	1499		0.93 (0.71, 1.21)
The Netherlands	1734		0.88 (0.62, 1.25)
UK - General population	3768		0.90 (0.74, 1.09)
UK - Health conscious	518		0.96 (0.58, 1.61)
ALL EPIC	18424		0.98 (0.89, 1.07)

Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Smoking related cancers were lung, bladder, upper aero-digestive, kidney, stomach, pancreatic, liver, colon, and rectum. Non-smoking related cancers (all other cancers).

3.4.5 Predicted 25-hydroxyvitamin D levels and digestive and non-digestive cancer incidences

Based on *a priori* hypothesis that cancers within the digestive system would be most responsive to vitamin D exposures, cases were split into digestive system and non-digestive cancers.

Digestive system cancers

The associations between predicted 25(OH)D and incidences of digestive system cancers are shown in Table 23. In the sexes combined multivariable model, a 25% reduced risk (95% CI: 0.66-0.85) of digestive system cancer incidence was observed when the highest and lowest quintiles of predicted circulating 25(OH)D were compared (*P*-trend <0.001). In the equivalent continuous model, a 37% lower digestive system cancer risk (95% CI: 0.52-0.76) was observed per 25 nmol/L higher predicted 25(OH)D. Similar risk estimates were observed after additional adjustment for physical activity index. The association observed amongst men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.65, 95% CI: 0.48-0.86) was slightly weaker than observed amongst women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.59, 95% CI: 0.46-0.76); however, this difference was not significant (*P*-interaction 0.35). Similar associations were observed when participants whose follow-up was less than 5 years were excluded from the analyses.

In sub-group analyses, consistent inverse associations were observed across the different levels of lifestyle, demographic, anthropometric, and dietary variables considered, with the exception of alcohol consumption, where no association was observed amongst participants who consumed 15-29.9 g/day (Table 24); although this interaction was non-significant when assessed statistically (*P*-interaction 0.21). The sexes combined associations for digestive system cancers were consistent across countries (*P*-heterogeneity 0.13), with only Spain not observing an inverse association (Figure 14A).

Non-digestive system cancers

For non-digestive system cancers, a weaker significant inverse association was observed in the multivariable models when men and women were analysed together (HR per 25 nmol/L increment in predicted 25(OH)D, 0.92, 95% CI: 0.85-0.99) (Table 23). This difference in the strength of associations with digestive system cancers was significant (*P*-heterogeneity <0.001). When men and women were analysed separately, similar strength non-significant risk estimates were observed in the continuous and categorical models (*P*-interaction 0.48).

In the sexes combined country specific analyses (Figure 14B), non-significant associations were observed across all constituent countries. Null or positive associations were observed for Denmark, France, Germany, and the UK (health conscious) cohorts. Despite this observed heterogeneity, these differences in risk estimates across countries were non-significant (*P*-heterogeneity 0.23).

Table 23. Risk (hazard ratios) of digestive and non-digestive cancer incidences amongst associated with predicted 25-hydroxyvitamin D level

	Quintile of predicted 25(OH)D (nmol/L)							<5 years excluded
	1	2	3	4	5		Per 25 nmol/L	Per 25 nmol/L
Men	<42.8	42.8-45.6	45.7-48.1	48.2-50.9	≥51.0		increase	increase
Women	<48.0	48.0-52.4	52.5-56.1	56.2-60.3	≥60.4	*P*-trend	increase	increase
Digestive system cancer incidence								
Both sexes								
N cases	1,168	1,001	821	821	814			
Basic model - HR (95% CI) †	1.00	0.86 (0.79-0.95)	0.74 (0.67-0.81)	0.77 (0.69-0.85)	0.67 (0.60-0.75)	<0.001	0.55 (0.48-0.65)	0.49 (0.41-0.59)
Multivariable model - HR (95% CI) ‡	1.00	0.91 (0.83-1.00)	0.80 (0.72-0.89)	0.84 (0.75-0.94)	0.75 (0.66-0.85)	<0.001	0.63 (0.52-0.76)	0.59 (0.48-0.73)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.92 (0.83-1.01)	0.81 (0.72-0.91)	0.85 (0.74-0.97)	0.74 (0.63-0.87)	<0.001	0.61 (0.48-0.77)	0.54 (0.41-0.72)
Men								
N cases	626	549	430	444	359			
Multivariable model - HR (95% CI) ‡	1.00	0.97 (0.86-1.10)	0.80 (0.69-0.92)	0.88 (0.76-1.02)	0.80 (0.68-0.94)	0.003	0.65 (0.48-0.86)	
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.94 (0.82-1.07)	0.76 (0.64-0.89)	0.80 (0.66-0.97)	0.69 (0.55-0.87)	0.001	0.46 (0.30-0.73)	
Women								
N cases	542	452	391	377	455			
Multivariable model - HR (95% CI) ‡	1.00	0.85 (0.74-0.98)	0.80 (0.68-0.94)	0.80 (0.67-0.96)	0.69 (0.56-0.85)	0.001	0.59 (0.46-0.76)	
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.88 (0.76-1.01)	0.85 (0.71-1.01)	0.86 (0.70-1.05)	0.75 (0.59-0.95)	0.03	0.63 (0.47-0.84)	
Non-digestive system cancers incidence								
Both sexes								
N cases	4,489	4,612	4,503	4,265	4,586			
Basic model - HR (95% CI) †	1.00	0.96 (0.92-1.00)	0.93 (0.89-0.97)	0.90 (0.86-0.94)	0.89 (0.84-0.93)	<0.001	0.86 (0.81-0.92)	0.85 (0.78-0.92)
Multivariable model - HR (95% CI) ‡	1.00	0.98 (0.94-1.03)	0.96 (0.91-1.01)	0.93 (0.88-0.98)	0.93 (0.87-0.99)	0.005	0.92 (0.85-0.99)	0.91 (0.82-1.00)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	1.00 (0.95-1.04)	0.98 (0.93-1.04)	0.96 (0.90-1.03)	0.97 (0.90-1.04)	0.27	0.99 (0.90-1.09)	0.93 (0.82-1.05)
Men								
N cases	1771	1569	1563	1369	1215			
Multivariable model - HR (95% CI) ‡	1.00	0.94 (0.88-1.01)	0.99 (0.91-1.07)	0.93 (0.85-1.01)	0.93 (0.85-1.02)	0.13	0.87 (0.74-1.02)	
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.97 (0.89-1.04)	1.03 (0.94-1.13)	0.98 (0.88-1.10)	1.02 (0.90-1.16)	0.69	1.03 (0.82-1.30)	
Women								
N cases	2718	3043	2940	2896	3371			
Multivariable model - HR (95% CI) ‡	1.00	1.00 (0.95-1.06)	0.94 (0.88-1.01)	0.93 (0.87-1.00)	0.92 (0.85-1.00)	0.019	0.93 (0.85-1.03)	
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	1.02 (0.96-1.08)	0.97 (0.90-1.04)	0.96 (0.89-1.04)	0.96 (0.87-1.05)	0.22	1.00 (0.89-1.11)	

† Basic model - Cox regression stratified by age (1-year categories), sex, and centre.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

ɸ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Digestive system cancers were stomach, colon and rectum, upper aero-digestive, liver, and pancreatic. Non-digestive system cancers (all other cancers).

Table 24. Risk (hazard ratios) of digestive system cancer incidence associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level across strata of lifestyle, demographic, anthropometric, and dietary variables

	Digestive system cancer incidence	
	Both sexes	P interaction
Stratification variable	Predicted 25(OH)D	
	HR (95% CI), per 25 nmol/L ‡	
Overall	0.63 (0.52-0.76)	
Sex		0.35
Men	0.65 (0.48-0.86)	
Women	0.59 (0.46-0.76)	
Age at recruitment		0.68
<55 years old	0.62 (0.45-0.85)	
55-<65 years old	0.64 (0.50-0.83)	
≥65 years old	0.63 (0.38-1.06)	
Follow-up		0.43
<5 years	0.80 (0.55-1.16)	
5-10 years	0.59 (0.45-0.78)	
≥10 years	0.68 (0.48-0.97)	
Smoking status		0.11
Never smoked	0.71 (0.53-0.96)	
Former smoker	0.53 (0.38-0.75)	
Current smoker	0.54 (0.39-0.75)	
Body mass index		0.19
<24.9 kg/m2	0.72 (0.54-0.97)	
<25.0-29.9 kg/m2	0.59 (0.44-0.80)	
≥30 kg/m2	0.52 (0.35-0.79)	
Physical activity		0.54
Inactive	0.68 (0.42-1.10)	
Moderately inactive	0.70 (0.46-1.06)	
Moderately active	0.51 (0.30-0.87)	
Active	0.59 (0.35-0.99)	
Alcohol consumption		0.21
Non-consumers	0.50 (0.30-0.84)	
<15 g/day	0.66 (0.51-0.85)	
15-29.9 g/day	1.02 (0.62-1.66)	
≥30 g/day	0.52 (0.34-0.80)	
Red and processed meat consumption *		0.31
Below median	0.63 (0.48-0.83)	
Above median	0.66 (0.51-0.84)	
Fruit and vegetable consumption *		0.19
Below median	0.68 (0.53-0.86)	
Above median	0.57 (0.43-0.76)	
Fish consumption *		0.93
Below median	0.56 (0.41-0.77)	
Above median	0.60 (0.48-0.77)	
Calcium intake *		0.33
Below median	0.64 (0.50-0.83)	
Above median	0.62 (0.47-0.81)	
Fibre intake *		0.52
Below median	0.71 (0.55-0.92)	
Above median	0.57 (0.43-0.74)	
Retinol intake *		0.83
Below median	0.66 (0.50-0.87)	
Above median	0.61 (0.47-0.78)	
Menopausal status		0.93
Premenopausal	0.51 (0.24-1.08)	
Postmenopausal	0.63 (0.47-0.84)	
Perimenopausal	0.36 (0.16-0.79)	
Surgical postmenopausal	0.92 (0.34-2.49)	
Ever use of contraceptive pill		0.72
No	0.59 (0.42-0.83)	
Yes	0.58 (0.40-0.85)	

* Median intakes: red and processed meat=93.3 g/day in men and 59.0 g/day in women; fruit and vegetable=356 g/day in men and 438 g/day in women; fish consumption=27.1 g/day in men and 21.9 g/day in women; calcium=972 mg/day in men and 952 mg/day in women; fibre=23.9 g/day
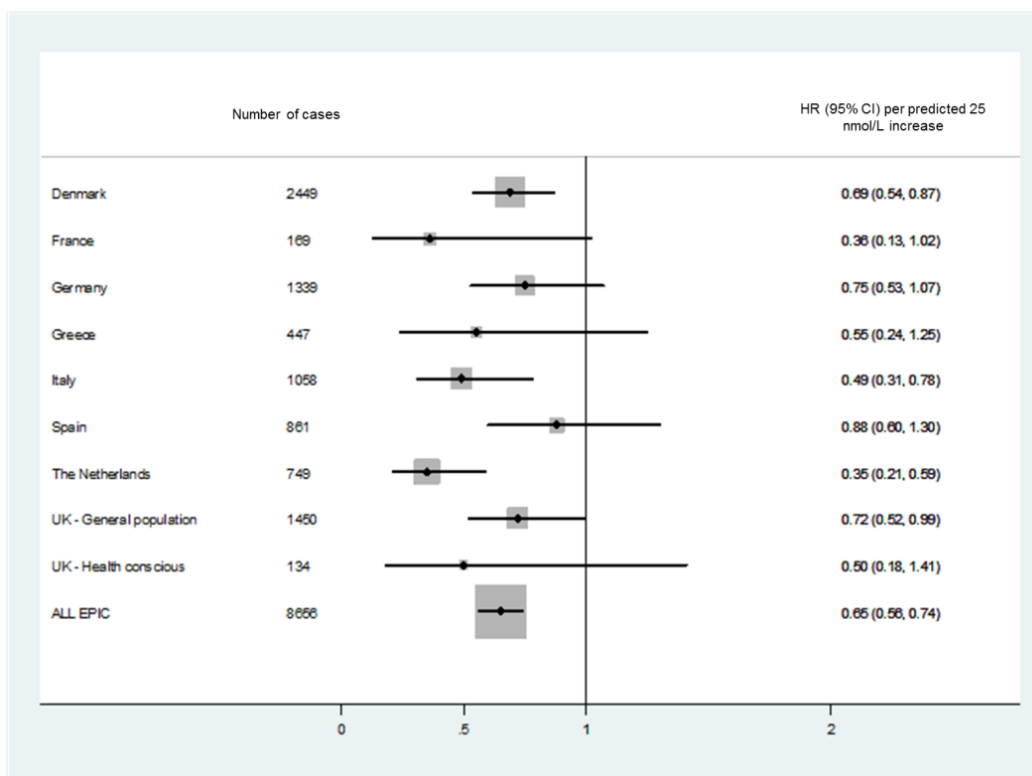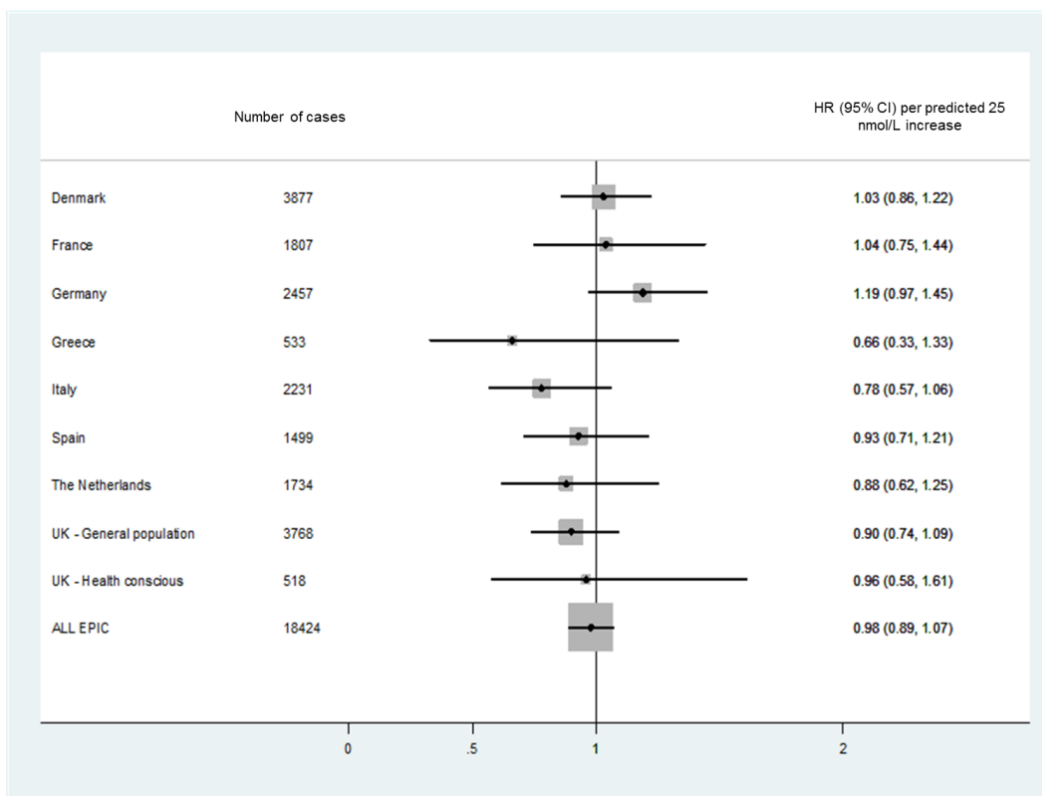
in men and 21.7 g/day in women; retinol=691 µg/day in men and 509 µg/day in women.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Digestive system cancers were stomach, colon and rectum, upper aero-digestive, liver, and pancreatic.

Figure 14. Risk (hazard ratios) of digestive system (A) and non-digestive system (B) cancer incidences (sexes combined), by country, associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level

(A)

(B)

Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former,

missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Digestive system cancers were stomach, colon and rectum, upper aero-digestive, liver, and pancreatic. Non-digestive system cancers (all other cancers).

3.4.6 Predicted 25-hydroxyvitamin D levels and incidence of individual cancers

Colorectal cancer

As was revealed in the validation stage of the predictor scores, significant inverse associations in the sexes combined categorical (n=2,112 cases; Q5 vs. Q1, HR 0.74, 95% CI: 0.61-0.90; *P*-trend 0.005) and continuous (HR per 25 nmol/L increment in predicted 25(OH)D, 0.75, 95% CI: 0.57-0.97) multivariable models were observed for colorectal cancer incidence. No significant heterogeneity was seen for the associations between predicted 25(OH)D with colon (n=1,345 cases; HR per 25 nmol/L increment in predicted 25(OH)D, 0.74, 95% CI: 0.53-1.04) and rectal cancers (n=767 cases; HR per 25 nmol/L increment in predicted 25(OH)D, 0.75, 95% CI: 0.54-1.05) (*P*-heterogeneity 0.83).The association was similar when participants whose follow-up time was less than 5 years were excluded from the analyses (Table 25). When men and women were analysed separately, similar inverse colorectal cancer risk estimates were observed in the highest quintiles of predicted 25(OH)D when compared against quintile 1 (a 25% reduced risk in men and a 28% reduced risk in women). In the continuous models, the inverse associations were slightly stronger amongst men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.68, 95% CI: 0.43-1.09) than in women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.74, 95% CI: 0.53-1.04) (Tables 26 and 27) (*P*-interaction 0.04).

The inverse colorectal cancer association was consistent across constituent countries, with inverse associations (HRs ranging from 0.21-0.76) observed in all countries except Spain, where a non-significant positive association was observed (*P*-heterogeneity 0.34).

Lung cancer

In the categorical sexes combined multivariable model (n=2,124 cases), a 24% reduced lung cancer incidence risk (95% CI: 0.63-0.92) was observed when the extreme quintiles of predicted circulating 25(OH)D were compared (*P*-trend 0.002). Inverse associations were observed in the equivalent continuous model (HR per 25 nmol/L increment in predicted

25(OH)D, 0.64, 95% CI: 0.49-0.84) (Table 25). The association was similar when participants whose follow-up time was less than 5 years were excluded from the analyses. This inverse association was stronger and only significant amongst women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.59, 95% CI: 0.41-0.85) and not men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.72, 95% CI: 0.48-1.08), and this difference was significant (*P*-interaction 0.004) (Tables 26 and 27). However, the inverse lung cancer incidence association was absent when the multivariable model was not adjusted for BMI (both sexes: HR per 25 nmol/L increment in predicted 25(OH)D, 1.01, 95% CI: 0.81-1.26) (Table A1 – appendix).

When the sexes combined association was analysed by smoking status, inverse associations of similar strength were observed amongst never smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.62, 95% CI: 0.26-1.52), former smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.68, 95% CI: 0.40-1.18) and current smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.66, 95% CI: 0.48-0.93). Overall the association across strata of smoking status was non-significant (*P*-interaction 0.45). Near significance heterogeneity was observed across countries with inverse associations observed across all countries, except France, Germany and Greece (*P*-heterogeneity 0.06).

Kidney cancer

During the follow-up period 623 kidney cancers accrued. A 42% reduced kidney cancer risk (95% CI: 0.40-0.83) was observed in the sexes combined multivariable model (*P*-trend 0.005) (Table 25). In the equivalent continuous model, a 53% lower kidney cancer risk (95% CI: 0.28-0.79) was observed per 25 nmol/L increment in predicted 25(OH)D. The association was similar when participants whose follow-up time was less than 5 years were excluded from the analyses. This association was stronger and significant amongst men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.33, 95% CI: 0.15-0.72) as compared to women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.64, 95% CI: 0.31-1.30) (*P*-interaction 0.03) (Tables 26 and 27).

Stomach and oesophageal cancers

Over the follow-up period, 862 stomach and oesophageal cancer cases were recorded. In the categorical multivariable model, a non-significant inverse association was observed (sexes combined: Q5 vs. Q1, HR 0.78, 95% CI: 0.58-1.05; *P*-trend 0.12) (Table 25). However, in the equivalent continuous model, a 40% lower risk (95% CI: 0.39-0.94) per 25

nmol/L of predicted 25(OH)D score was observed. An inverse association of similar magnitude (albeit non-significant) was observed when participants whose follow-up time was less than 5 years were excluded from the analyses. When men and women were analysed separately, non-significant inverse associations were observed (*P*-interaction 0.66) (Tables 26 and 27). However, when the sexes combined multivariable model was not adjusted for BMI, the inverse association attenuated and was no longer significant (HR per 25 nmol/L increment in predicted 25(OH)D, 0.74, 95% CI: 0.51-1.07) (Table A1 – appendix).

Bladder cancer

Non-significant associations were observed for bladder cancer incidence (n=363 cases) when men and women were analysed collectively (multivariable model: Q5 vs. Q1, HR 0.87, 95% CI: 0.69-1.11; *P*-trend 0.59) (Table 25). When analysed separately, similar associations were observed (*P*-interaction 0.97) (Tables 26 and 27).

Pancreatic cancer

Inverse associations for pancreatic cancer incidence (n=337 cases) were observed in the sexes combined continuous multivariable model (HR per 25 nmol/L increment in predicted 25(OH)D, 0.52, 95% CI: 0.32-0.84) (Table 25). The inverse association was slightly stronger when participants whose follow-up time was less than 5 years were excluded from the analyses. The association was significant only for women, where a significant 58% lower risk (95% CI: 0.22-0.80) was observed per 25 nmol/L of predicted 25(OH)D score; amongst men, a non-significant 30% lower risk was observed for a same increment in predicted 25(OH)D score (Tables 26 and 27). This difference in association between the sexes was non-significant (*P*-interaction 0.78). Non-significant inverse associations were observed in all of the categorical models.

Liver cancer

Non-significant inverse associations were observed in the sexes combined multivariable model for liver cancer incidence (n=205 cases; Q5 vs. Q1, HR 0.80, 95% CI: 0.52-1.24; *P*-trend 0.17) (Table 25). Non-significant associations were also observed when men and women were analysed separately (*P*-interaction 0.78) (Tables 26 and 27). When the sexes combined multivariable model was not adjusted for BMI, the inverse association strengthened and became significant (HR per 25 nmol/L increment in predicted 25(OH)D, 0.52, 95% CI: 0.31-0.87) (Table A1 – appendix).

Non-Hodgkin's lymphoma

Non-significant associations were observed for non-Hodgkin's lymphoma incidence (n=626 cases) in the sexes combined multivariable model (Q5 vs. Q1, HR 0.79, 95% CI: 0.56-1.12; *P*-trend 0.25) (Table 25). Similar non-significant associations were also observed for men and women (*P*-interaction 0.35) (Tables 26 and 27).

Brain cancer

In the sexes combined categorical multivariable model, a near significance 25% reduced risk (95% CI: 0.54-1.04) of brain cancer incidence (n=764 cases) was observed amongst those in the highest predicted 25(OH)D quintile when compared versus those in the lowest quintile (*P*-trend 0.07) (Table 25). A non-significant inverse association was also observed in the equivalent continuous model. The inverse association, although once more non-significant, was more apparent amongst men (Q5 vs. Q1, HR 0.63, 95% CI: 0.40-1.01; *P*-trend 0.05) as compared to women (Q5 vs. Q1, HR 0.84, 95% CI: 0.52-1.36; *P*-trend 0.38); however, this difference was non-significant (*P*-interaction 0.72) (Tables 26 and 27).

Skin cancer

Predicted circulating 25(OH)D was positively associated with the risk of skin cancer incidence (n=1,550 cases) in the sexes combined multivariable model (Q5 vs. Q1, HR 1.25, 95% CI: 1.03-1.51; *P*-trend 0.02) (Table 25). In the continuous models, a 26% higher skin cancer risk (95% CI: 1.00-1.60) was associated with higher predicted 25(OH)D, although this association did not reach the significance threshold. Higher skin cancer risks were also observed when the participants whose follow-up time was less than 5 years were excluded from the analyses. Positive associations were observed amongst men and women (*P*-interaction 0.26); although, in women these associations were stronger and significant (Q5 vs. Q1, HR 1.40, 95% CI: 1.08-1.83; *P*-trend 0.01) (Tables 26 and 27).

Thyroid cancer

In the sexes combined multivariable model, a 50% lower risk of thyroid cancer incidence (95% CI: 0.26-0.97) was observed per 25 nmol/L of predicted 25(OH)D (n=325 cases) (Table 25). Similar inverse associations were observed when the first 5 years of follow-up was excluded from the analyses. In the equivalent categorical model a non-significant inverse association was observed (Q5 vs. Q1, HR 0.72, 95% CI: 0.43-1.21; *P*-trend 0.20).

Similar strength, albeit non-significant, inverse associations were observed for men and women when analysed separately (*P*-interaction 0.31) (Tables 26 and 27). When the sexes combined multivariable model was not adjusted for BMI, the inverse association weakened and became non-significant (HR per 25 nmol/L increment in predicted 25(OH)D, 0.69, 95% CI: 0.41-1.15) (Table A1 – appendix).

Prostate cancer

As was revealed in the validation stages of the predictor scores, a non-significant positive association was observed for risk of prostate cancer incidence (n=2,553 cases) in the multivariable models (Q5 vs. Q1, 1.11, 95% CI: 0.95-1.30; *P*-trend 0.51) (Table 26).

Breast, ovarian and endometrial cancers

Null associations were observed for risk of breast cancer (n=7,144 cases: Q5 vs. Q1, 0.97, 95% CI: 0.86-1.10; *P*-trend 0.33) and ovarian cancer (n=811 cases: Q5 vs. Q1, 1.00, 95% CI: 0.70-1.42; *P*-trend 0.93) incidences in the multivariable models (Table 27).

For endometrial cancer incidence (n=893 cases), a non-significant inverse association was observed when the extreme quintiles of predicted circulating 25(OH)D were compared (Q5 vs. Q1, HR 0.85, 95% CI: 0.61-1.19; *P*-trend 0.31) (Table 27). For the equivalent continuous model a non-significant 8% lower endometrial cancer incidence risk (95% CI: 0.63-1.35) was observed per 25 nmol/L increment in predicted 25(OH)D. However, when the multivariable model was not adjusted for BMI, the inverse association strengthened and became significant (HR per 25 nmol/L increment in predicted 25(OH)D, 0.36, 95% CI: 0.27-0.49) (Table A1 – appendix).

Table 25. Risk (hazard ratios) for incidence of specific cancer types amongst men and women by predicted 25-hydroxyvitamin D categories

Both sexes	Quintile of predicted 25(OH)D (nmol/L)							<5 years excluded
	1	2	3	4	5		Per 25 nmol/L increase	Per 25 nmol/L increase
Men	<42.8	42.8-45.6	45.7-48.1	48.2-50.9	≥51.0			
Women	<48.0	48.0-52.4	52.5-56.1	56.2-60.3	≥60.4	P-trend		
Colorectal cancer								
N cases	517	426	396	392	381			
Multivariable model - HR (95% CI) † ¥	1.00	0.82 (0.71-0.94)	0.80 (0.68-0.93)	0.84 (0.71-0.99)	0.74 (0.61-0.90)	0.005	0.75 (0.57-0.97)	0.65 (0.49-0.88)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.84 (0.72-0.97)	0.83 (0.70-0.99)	0.88 (0.72-1.07)	0.77 (0.61-0.97)	0.06	0.83 (0.59-1.15)	0.71 (0.49-1.03)
Lung cancer								
N cases	473	415	376	350	510			
Multivariable model - HR (95% CI) †	1.00	0.92 (0.80-1.07)	0.85 (0.73-1.00)	0.79 (0.66-0.93)	0.76 (0.63-0.92)	0.002	0.64 (0.49-0.84)	0.63 (0.45-0.87)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.97 (0.83-1.13)	0.93 (0.77-1.11)	0.87 (0.71-1.07)	0.87 (0.69-1.10)	0.19	0.77 (0.55-1.08)	0.77 (0.51-1.15)
Kidney cancer								
N cases	185	132	131	89	86			
Multivariable model - HR (95% CI) †	1.00	0.83 (0.65-1.06)	0.93 (0.71-1.22)	0.69 (0.50-0.94)	0.58 (0.40-0.83)	0.005	0.47 (0.28-0.79)	0.49 (0.26-0.93)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.83 (0.64-1.08)	0.94 (0.69-1.28)	0.70 (0.48-1.02)	0.61 (0.39-0.95)	0.04	0.47 (0.24-0.92)	0.43 (0.19-0.99)
Stomach & oesophageal cancer								
N cases	215	191	165	148	143			
Multivariable model - HR (95% CI) †	1.00	0.97 (0.78-1.20)	0.94 (0.74-1.20)	0.89 (0.69-1.16)	0.78 (0.58-1.05)	0.12	0.60 (0.39-0.94)	0.65 (0.63-1.18)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.98 (0.78-1.23)	0.96 (0.73-1.26)	0.90 (0.66-1.23)	0.76 (0.52-1.11)	0.19	0.53 (0.30-0.94)	0.49 (0.24-0.98)
Bladder cancer								
N cases	77	82	64	62	78			
Multivariable model - HR (95% CI) †	1.00	0.91 (0.76-1.09)	0.97 (0.80-1.18)	1.04 (0.84-1.27)	0.87 (0.69-1.11)	0.59	0.89 (0.62-1.28)	1.01 (0.66-1.55)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.96 (0.79-1.17)	1.07 (0.85-1.34)	1.17 (0.91-1.51)	1.02 (0.75-1.38)	0.56	1.11 (0.70-1.75)	1.29 (0.77-2.17)
Pancreatic cancer								
N cases	82	69	51	60	75			
Multivariable model - HR (95% CI) †	1.00	1.02 (0.79-1.31)	0.71 (0.53-0.95)	0.90 (0.66-1.22)	0.77 (0.55-1.09)	0.11	0.52 (0.32-0.84)	0.39 (0.22-0.70)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	1.06 (0.81-1.38)	0.75 (0.54-1.05)	0.97 (0.67-1.39)	0.83 (0.54-1.28)	0.39	0.50 (0.27-0.94)	0.34 (0.16-0.72)
Liver cancer								
N cases	74	36	34	27	34			
Multivariable model - HR (95% CI) †	1.00	0.85 (0.64-1.13)	0.71 (0.50-0.99)	0.79 (0.54-1.15)	0.80 (0.52-1.24)	0.17	0.80 (0.41-1.43)	0.77 (0.36-1.65)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.82 (0.61-1.12)	0.66 (0.45-0.98)	0.71 (0.46-1.11)	0.68 (0.40-1.16)	0.09	0.69 (0.31-1.51)	0.66 (0.25-1.73)
Non-Hodgkin's lymphoma								
N cases	133	113	127	117	136			
Multivariable model - HR (95% CI) †	1.00	0.76 (0.58-1.00)	0.87 (0.65-1.16)	0.79 (0.58-1.09)	0.79 (0.56-1.12)	0.25	0.87 (0.54-1.43)	0.95 (0.56-1.59)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.73 (0.55-0.98)	0.81 (0.59-1.13)	0.72 (0.50-1.05)	0.71 (0.46-1.09)	0.14	0.81 (0.44-1.48)	0.76 (0.40-1.46)
Brain cancer								
N cases	185	178	153	131	117			
Multivariable model - HR (95% CI) †	1.00	0.99 (0.79-1.25)	0.92 (0.71-1.19)	0.83 (0.62-1.11)	0.75 (0.54-1.04)	0.07	0.77 (0.49-1.22)	0.82 (0.46-1.46)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	1.05 (0.82-1.33)	1.00 (0.74-1.33)	0.92 (0.66-1.29)	0.84 (0.56-1.25)	0.39	1.01 (0.59-1.72)	0.91 (0.45-1.83)
Skin cancer								
N cases	181	298	344	392	335			
Multivariable model - HR (95% CI) †	1.00	1.12 (0.96-1.31)	1.16 (0.99-1.37)	1.22 (1.03-1.44)	1.25 (1.03-1.51)	0.02	1.26 (1.00-1.60)	1.18 (0.87-1.60)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	1.11 (0.95-1.31)	1.15 (0.96-1.38)	1.20 (0.99-1.47)	1.23 (0.98-1.55)	0.07	1.20 (0.89-1.61)	1.03 (0.70-1.50)
Thyroid cancer								
N cases	90	83	70	49	33			
Multivariable model - HR (95% CI) †	1.00	0.90 (0.66-1.24)	0.83 (0.58-1.20)	0.81 (0.53-1.23)	0.72 (0.43-1.21)	0.20	0.50 (0.26-0.97)	0.52 (0.23-1.18)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.98 (0.70-1.37)	0.96 (0.63-1.44)	0.95 (0.59-1.55)	0.88 (0.49-1.60)	0.72	0.61 (0.27-1.36)	0.64 (0.23-1.72)

† Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m²), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

¥ Multivariable model – plus adjustment for cereal fibre intake (g/day).

ɸ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Table 26. Risk (hazard ratios) for incidence of specific cancer types amongst men by predicted 25-hydroxyvitamin D categories

Men	Quintile of predicted 25(OH)D (nmol/L)					P-trend	Per 25 nmol/L increase
	1	2	3	4	5		
	<42.8	42.8-45.6	45.7-48.1	48.2-50.9	≥51.0		
Colorectal cancer							
N cases	251	187	178	182	145		
Multivariable model - HR (95% CI) † ¥	1.00	0.78 (0.63-0.95)	0.75 (0.60-0.94)	0.84 (0.67-1.06)	0.75 (0.58-0.97)	0.08	0.68 (0.43-1.09)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	0.80 (0.64-0.99)	0.79 (0.61-1.03)	0.88 (0.65-1.19)	0.75 (0.52-1.07)	0.24	0.72 (0.36-1.44)
Lung cancer							
N cases	325	252	254	197	177		
Multivariable model - HR (95% CI) †	1.00	0.89 (0.75-1.07)	0.99 (0.82-1.20)	0.83 (0.67-1.03)	0.87 (0.69-1.09)	0.19	0.72 (0.48-1.08)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	0.98 (0.81-1.19)	1.16 (0.92-1.46)	1.01 (0.76-1.32)	1.07 (0.78-1.48)	0.66	1.05 (0.59-1.86)
Kidney cancer							
N cases	112	83	71	57	32		
Multivariable model - HR (95% CI) †	1.00	0.89 (0.65-1.21)	0.84 (0.59-1.18)	0.74 (0.50-1.09)	0.49 (0.31-0.79)	0.004	0.33 (0.15-0.72)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	0.92 (0.66-1.29)	0.89 (0.59-1.35)	0.83 (0.51-1.35)	0.60 (0.33-1.12)	0.15	0.43 (0.13-1.43)
Stomach & oesophageal cancer							
N cases	127	129	98	96	75		
Multivariable model - HR (95% CI) †	1.00	1.11 (0.85-1.44)	0.93 (0.68-1.25)	0.95 (0.69-1.31)	0.84 (0.59-1.20)	0.24	0.67 (0.36-1.24)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	1.07 (0.80-1.42)	0.87 (0.61-1.25)	0.87 (0.57-1.31)	0.74 (0.45-1.20)	0.16	0.47 (0.18-1.21)
Bladder cancer							
N cases	220	174	192	189	138		
Multivariable model - HR (95% CI) †	1.00	0.82 (0.66-1.02)	0.93 (0.74-1.16)	1.02 (0.80-1.28)	0.88 (0.68-1.15)	0.87	0.87 (0.55-1.39)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	0.86 (0.68-1.09)	1.01 (0.77-1.32)	1.13 (0.83-1.54)	1.02 (0.71-1.47)	0.49	1.04 (0.54-2.02)
Pancreatic cancer							
N cases	74	78	48	62	47		
Multivariable model - HR (95% CI) †	1.00	1.20 (0.85-1.69)	0.74 (0.49-1.11)	0.98 (0.65-1.48)	0.85 (0.54-1.35)	0.34	0.70 (0.32-1.56)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	1.05 (0.72-1.53)	0.60 (0.37-0.97)	0.76 (0.44-1.30)	0.62 (0.33-1.17)	0.09	0.34 (0.10-1.20)
Liver cancer							
N cases	82	58	29	33	29		
Multivariable model - HR (95% CI) †	1.00	1.03 (0.71-1.48)	0.62 (0.39-1.00)	0.81 (0.50-1.33)	0.94 (0.55-1.62)	0.47	0.74 (0.28-1.93)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	0.94 (0.62-1.43)	0.54 (0.31-0.95)	0.65 (0.34-1.23)	0.69 (0.33-1.45)	0.20	0.42 (0.09-1.90)
Non-Hodgkin's lymphoma							
N cases	64	54	60	53	53		
Multivariable model - HR (95% CI) †	1.00	0.80 (0.54-1.19)	0.89 (0.59-1.33)	0.77 (0.49-1.20)	0.84 (0.53-1.35)	0.52	0.83 (0.36-1.94)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	0.72 (0.47-1.10)	0.73 (0.45-1.20)	0.60 (0.34-1.06)	0.63 (0.33-1.22)	0.20	0.51 (0.14-1.89)
Brain cancer							
N cases	75	66	70	54	46		
Multivariable model - HR (95% CI) †	1.00	0.86 (0.60-1.22)	0.92 (0.63-1.35)	0.73 (0.48-1.11)	0.63 (0.40-1.01)	0.05	0.47 (0.20-1.07)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	0.93 (0.63-1.36)	1.06 (0.67-1.67)	0.86 (0.50-1.48)	0.76 (0.40-1.44)	0.41	0.74 (0.22-2.48)
Skin cancer							
N cases	148	165	188	170	160		
Multivariable model - HR (95% CI) †	1.00	1.03 (0.81-1.30)	1.17 (0.92-1.50)	1.10 (0.85-1.43)	1.09 (0.82-1.44)	0.52	1.19 (0.74-1.91)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	1.01 (0.79-1.30)	1.14 (0.86-1.52)	1.08 (0.77-1.50)	1.08 (0.74-1.58)	0.70	1.19 (0.59-2.38)
Thyroid cancer							
N cases	11	11	10	14	9		
Multivariable model - HR (95% CI) †	1.00	0.87 (0.33-2.14)	0.71 (0.26-1.91)	1.07 (0.40-2.90)	0.83 (0.27-2.59)	0.92	0.39 (0.05-3.08)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	0.93 (0.35-2.48)	0.80 (0.24-2.66)	1.35 (0.37-4.97)	1.24 (0.26-5.900)	0.64	0.30 (0.01-8.52)
Prostate cancer							
N cases	581	568	522	448	434		
Multivariable model - HR (95% CI) †	1.00	1.09 (0.96-1.23)	1.06 (0.92-1.21)	0.98 (0.85-1.14)	1.11 (0.95-1.30)	0.51	1.12 (0.86-1.47)
Multivariable model + physical activity adj - HR (95% CI) ϕ	1.00	1.10 (0.96-1.25)	1.08 (0.92-1.27)	1.02 (0.84-1.23)	1.19 (0.96-1.47)	0.24	1.30 (0.91-1.85)

† Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories) and centre.

¥ Multivariable model – plus adjustment for cereal fibre intake (g/day).

ɸ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Table 27. Risk (hazard ratios) for incidence of specific cancer types amongst women by predicted 25-hydroxyvitamin D categories

Women	1	2	3	4	5	P-trend	Per 25 nmol/L increase
	<48.0	48.0-52.4	52.5-56.1	56.2-60.3	≥60.4		
Colorectal cancer							
N cases	266	239	218	210	236		
Multivariable model - HR (95% CI) † ¥	1.00	0.86 (0.71-1.05)	0.84 (0.67-1.05)	0.84 (0.66-1.07)	0.72 (0.54-0.97)	0.05	0.74 (0.53-1.04)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.88 (0.71-1.07)	0.87 (0.68-1.10)	0.87 (0.66-1.14)	0.76 (0.55-1.06)	0.15	0.80 (0.54-1.18)
Lung cancer							
N cases	148	163	122	153	333		
Multivariable model - HR (95% CI) †	1.00	0.91 (0.70-1.17)	0.59 (0.44-0.80)	0.61 (0.45-0.83)	0.53 (0.38-0.76)	<0.001	0.59 (0.41-0.85)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.90 (0.69-1.17)	0.59 (0.43-0.81)	0.61 (0.44-0.85)	0.54 (0.36-0.80)	0.001	0.64 (0.42-0.98)
Kidney cancer							
N cases	73	49	60	32	54		
Multivariable model - HR (95% CI) †	1.00	0.77 (0.51-1.15)	1.13 (0.73-1.75)	0.65 (0.38-1.13)	0.76 (0.41-1.42)	0.40	0.64 (0.31-1.30)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.74 (0.48-1.13)	1.07 (0.66-1.73)	0.61 (0.33-1.11)	0.69 (0.34-1.40)	0.30	0.51 (0.22-1.19)
Stomach & oesophageal cancer							
N cases	88	62	67	52	68		
Multivariable model - HR (95% CI) †	1.00	0.77 (0.53-1.11)	0.99 (0.66-1.48)	0.82 (0.51-1.30)	0.70 (0.40-1.21)	0.29	0.55 (0.29-1.04)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.82 (0.56-1.20)	1.10 (0.71-1.70)	0.92 (0.55-1.54)	0.80 (0.42-1.49)	0.59	0.57 (0.27-1.21)
Bladder cancer							
N cases	77	82	64	62	78		
Multivariable model - HR (95% CI) †	1.00	1.12 (0.79-1.59)	1.05 (0.70-1.57)	1.03 (0.66-1.62)	0.77 (0.45-1.32)	0.36	0.80 (0.44-1.45)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.20 (0.84-1.72)	1.17 (0.76-1.82)	1.19 (0.72-1.96)	0.93 (0.51-1.70)	0.79	1.00 (0.50-1.98)
Pancreatic cancer							
N cases	82	69	51	60	75		
Multivariable model - HR (95% CI) †	1.00	0.85 (0.59-1.24)	0.67 (0.44-1.04)	0.82 (0.52-1.30)	0.70 (0.40-1.20)	0.22	0.42 (0.22-0.80)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.97 (0.66-1.43)	0.84 (0.52-1.34)	1.08 (0.64-1.80)	0.98 (0.53-1.82)	0.95	0.56 (0.27-1.19)
Liver cancer							
N cases	74	36	34	27	34		
Multivariable model - HR (95% CI) †	1.00	0.66 (0.42-1.03)	0.78 (0.47-1.310)	0.76 (0.42-1.37)	0.64 (0.31-1.33)	0.25	0.55 (0.24-1.28)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.66 (0.41-1.05)	0.79 (0.45-1.37)	0.76 (0.39-1.46)	0.63 (0.28-1.42)	0.32	0.52 (0.19-1.40)
Non-Hodgkin's lymphoma							
N cases	69	59	67	64	83		
Multivariable model - HR (95% CI) †	1.00	0.74 (0.50-1.09)	0.86 (0.57-1.31)	0.82 (0.52-1.30)	0.74 (0.43-1.27)	0.41	0.87 (0.47-1.60)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.73 (0.49-1.10)	0.85 (0.54-1.34)	0.81 (0.49-1.36)	0.73 (0.40-1.34)	0.43	0.88 (0.43-1.79)
Brain cancer							
N cases	110	112	83	77	71		
Multivariable model - HR (95% CI) †	1.00	1.11 (0.82-1.50)	0.91 (0.64-1.31)	0.92 (0.62-1.38)	0.84 (0.52-1.36)	0.38	0.98 (0.58-1.67)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.15 (0.84-1.58)	0.97 (0.66-1.43)	0.99 (0.64-1.56)	0.92 (0.54-1.58)	0.65	1.15 (0.68-1.92)
Skin cancer							
N cases	181	298	344	392	335		
Multivariable model - HR (95% CI) †	1.00	1.19 (0.97-1.46)	1.18 (0.95-1.46)	1.31 (1.04-1.64)	1.40 (1.08-1.83)	0.01	1.26 (0.95-1.66)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.19 (0.96-1.47)	1.18 (0.93-1.49)	1.31 (1.01-1.69)	1.39 (1.03-1.88)	0.03	1.19 (0.85-1.66)
Thyroid cancer							
N cases	90	83	70	49	33		
Multivariable model - HR (95% CI) †	1.00	0.92 (0.65-1.29)	0.86 (0.58-1.28)	0.76 (0.48-1.22)	0.70 (0.39-1.26)	0.19	0.55 (0.27-1.11)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.00 (0.70-1.44)	0.99 (0.64-1.54)	0.90 (0.53-1.53)	0.86 (0.44-1.67)	0.63	0.71 (0.31-1.63)
Breast cancer							
N cases	1,234	1,485	1,440	1,364	1,621		
Multivariable model - HR (95% CI) † ¶	1.00	1.04 (0.96-1.14)	0.98 (0.89-1.08)	0.95 (0.86-1.06)	0.97 (0.86-1.10)	0.33	1.02 (0.89-1.17)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.06 (0.97-1.16)	1.01 (0.91-1.12)	0.99 (0.88-1.11)	1.02 (0.89-1.17)	0.87	1.12 (0.96-1.31)
Ovarian cancer							
N cases	169	164	151	157	170		
Multivariable model - HR (95% CI) † ¶	1.00	1.03 (0.80-1.31)	0.99 (0.75-1.31)	1.10 (0.81-1.48)	1.00 (0.70-1.42)	0.93	1.25 (0.85-1.85)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.99 (0.77-1.28)	0.94 (0.70-1.27)	1.03 (0.74-1.44)	0.92 (0.62-1.38)	0.75	1.22 (0.77-1.93)
Endometrial cancer							
N cases	221	186	163	141	182		
Multivariable model - HR (95% CI) † ¶	1.00	0.94 (0.75-1.17)	0.92 (0.72-1.19)	0.86 (0.65-1.14)	0.85 (0.61-1.19)	0.31	0.92 (0.63-1.35)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.91 (0.72-1.14)	0.88 (0.67-1.15)	0.80 (0.59-1.10)	0.77 (0.53-1.12)	0.16	0.83 (0.53-1.30)

† Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories) and centre.

¥ Multivariable model – plus adjustment for cereal fibre intake (g/day).

¶ Multivariable model – plus adjustment for age at first pregnancy (<21; 21-<30; 30+ years old; no children; or not specified); and age at menarche (<12; 12-<15; 15+ years old; or not specified).

ϕ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

3.5 Application of the predicted 25-hydroxyvitamin D score within the European Prospective Investigation into Cancer and Nutrition to assess mortality risk

3.5.1 Characteristics of cohort participants

The cohort characteristics by country and sex are shown in Table 28. The highest predicted circulating 25(OH)D levels were observed in the Netherlands for men and Denmark for women. The lowest predicted circulating 25(OH)D levels for men and women were observed amongst participants from Germany and Greece respectively. After adjustment for age, using the European standard population (175), the all-cause mortality rates for men and women were 94 and 60 cases per 10,000 person-years respectively.

Table 28. Cohort characteristics by sex and country of participants included in the predicted 25(OH)D-all-cause mortality analyses

Country	*N* of participants		Total person-years		*N* all-cause deaths		Mean (SD) predicted 25(OH)D level (nmol/L)	
	Men	Women	Men	Women	Men	Women	Men	Women
Denmark	23,143	26,594	268,367	314,974	2,305	1,692	48.1 (4.0)	65.2 (6.7)
France	-	18,518	-	279,154	-	618	-	54.5 (4.7)
Germany	18,622	25,922	209,411	294,134	1,103	629	42.8 (4.1)	49.3 (5.8)
Greece	8,768	12,931	82,210	128,805	718	505	45.1 (3.7)	44.5 (5.2)
Italy	12,891	29,001	162,061	352,568	522	757	48.6 (3.7)	49.2 (5.0)
Spain	13,599	23,057	183,829	315,870	848	566	46.8 (4.5)	56.5 (6.8)
The Netherlands	7,068	22,667	89,267	292,321	277	1,192	50.7 (4.0)	56.4 (4.9)
United Kingdom	19,160	48,816	247,064	630,676	1,965	2,564	49.0 (4.2)	55.8 (5.2)
All EPIC	103,251	207,506	1,242,208	2,608,501	7,738	8,523	47.1 (4.7)	54.6 (7.8)

3.5.2 Baseline characteristics by predicted 25-hydroxyvitamin D quintiles

The characteristics of the participants included in the analyses of mortality by quintiles of predicted circulating 25(OH)D are shown in Table 29. Compared to participants in the lowest quintile, those participants in the highest quintile of predicted 25(OH)D were more likely to be younger, physically active and have a lower BMI. Men in the highest predicted 25(OH)D quintile had lower intakes of red and processed meat and higher intakes of total energy and fruits and vegetables than those in the lowest quintile. Conversely, women in the highest

predicted 25(OH)D quintile had higher red and processed meat intakes and lower intakes of total energy and fruits and vegetables than those in the lowest quintile. Women in the highest predicted 25(OH)D quintile were more likely to be current smokers than those in the lowest quintile; whereas amongst men, current smokers were relatively evenly distributed across the predicted 25(OH)D categories.

Table 29. Baseline characteristics of study participants included in the all-cause mortality analyses by categories (quintiles) of predicted 25-hydroxyvitamin D score

Characteristic	Quintile of predicted 25(OH)D				
	Q1	Q2	Q3	Q4	Q5
Men					
Predicted 25(OH)D range (g/day)	<43.0	43.0-45.9	46.0-48.3	48.4-51.1	≥51.2
N	20,651	20,650	20,650	20,650	20,650
***N* all-cause deaths**	2155	1784	1476	1286	1037
Age at recruitment (years)	54.2 (8.7)	52.9 (9.0)	51.7 (9.2)	50.5 (9.5)	48.2 (10.7)
Body mass index (kg/m2)	29.1 (3.9)	27.4 (3.4)	26.5 (3.1)	25.6 (2.9)	24.3 (2.7)
Education ‡ Longer education including University (%)	30.3	28.5	28.3	29.0	27.9
Smoking status and intensity ‡					
Current (%)	31.1	32.2	31.7	31.5	30.6
Physical activity ‡					
Active (%)	3.3	10.6	18.1	34.3	69.1
Total energy intake (kcal/day)	2371 (671)	2414 (662)	2442 (641)	2466 (644)	2523 (662)
Red and processed meat consumption (g/day)	109.3 (64.4)	101.7 (61.7)	98.2 (61.1)	96.8 (61.7)	94.6 (63.8)
Fruit & vegetable consumption (g/day)	384.9 (296.3)	435.4 (307.7)	445.9 (291.0)	438.5 (274.3)	430.0 (264.5)
Vitamin D intake (µg/day)	3.3 (1.8)	3.7 (2.1)	3.9 (2.2)	4.3 (2.4)	5.5 (3.8)
Calcium intake (mg/day)	945.3 (405)	1005.7 (415)	1045 (415)	1078 (423)	1126 (439)
Alcohol intake (g/day)	25.1 (27.7)	23.7 (25.0)	23.4 (23.6)	22.5 (22.3)	19.8 (20.1)
Women					
Predicted 25(OH)D range (g/day)	<48.3	48.3-52.6	52.7-56.2	56.3-60.4	≥60.5
N	41,503	41,500	41,501	41,501	41,501
***N* all-cause deaths**	2180	1762	1473	1365	1743
Age at recruitment (years)	53.4 (10.1)	50.5 (10.2)	48.8 (10.6)	47.6 (11.3)	51.2 (9.8)
Body mass index (kg/m2)	29.6 (5.1)	25.9 (3.9)	24.4 (3.7)	23.6 (3.5)	23.4 (3.1)
Education ‡ Longer education including University (%)	13.4	22.1	27.5	29.4	19.6
Smoking status and intensity ‡					
Current (%)	17.9	18.8	17.4	18.0	26
Physical activity ‡					
Active (%)	4.5	9.9	15.2	23.4	34.5
Total energy intake (kcal/day)	1970 (572)	1968 (550)	1958 (534)	1936 (509)	1921 (481)
Red and processed meat consumption (g/day)	66.4 (42.8)	64.6 (43.2)	60.7 (44.9)	57.1 (44.7)	66.9 (40.3)
Fruit & vegetable consumption (g/day)	527.8 (309.6)	482.2 (283.8)	485.2 (263.0)	486.9 (263.4)	450.0 (245.4)
Vitamin D intake (µg/day)	2.5 (1.5)	2.8 (1.5)	3.0 (1.7)	3.3 (1.8)	4.5 (2.6)
Calcium intake (mg/day)	952.0 (377)	995.2 (396)	1019.7 (392)	1031.6 (385)	1039.7 (408)
Alcohol intake (g/day)	6.8 (11.5)	8.2 (11.9)	8.7 (11.9)	8.7 (11.9)	10.5 (12.6)
Ever use of contraceptive pill ‡					
Yes (%)	39.6	56.0	64.7	69.0	64.3
Ever use of menopausal hormone therapy ‡					
Yes (%)	13.4	18.5	19.6	23.8	39.6
Menopausal status ‡					
Postmenopausal (%)	51.9	41.2	36.0	35.6	51.7

Mean and standard deviation (in parenthesis) for continuous variables, or percentages for categorical variables (‡).

3.5.3 Predicted 25-hydroxyvitamin D levels and all-cause mortality risk

In the sexes combined categorical model, higher levels of predicted 25(OH)D were associated with a reduced all-cause mortality (Q5 vs. Q1, HR 0.65, 95% CI: 0.61-0.70; *P*-trend <0.001) (Table 30). In the equivalent continuous model, a 52% lower all-cause mortality risk (95% CI: 0.43-0.53) was observed per 25 nmol/L increment of predicted 25(OH)D. Similar associations were observed when participants whose follow-up times were less than 5 years were excluded from the analyses. When analysed by sex, similar inverse associations were observed for men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.46, 95% CI: 0.39-0.54) and women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.47, 95% CI: 0.41-0.53) were observed in the multivariable models (*P*-interaction 0.84).

Significant predicted 25(OH)D-all-cause mortality associations were observed across strata of all lifestyle, demographic, anthropometric, and dietary variables considered (Table 31). For BMI, a significant interaction was observed (*P*-interaction 0.001); however, significant associations were observed across all categories of BMI.

The mutually adjusted HRs for all-cause mortality associated with the individual components of the applied predictor scores are shown in Table 32. After multivariable adjustment, waist circumference, country, and physical activity index, were significantly associated with risk of all-cause death. In contrast, dietary vitamin D intake and ever use of menopausal hormones were not significantly associated with all-cause mortality risk.

Significant inverse associations were observed across all countries; although the strength of the relationships varied (*P*-heterogeneity <0.001), with the strongest association observed for the Netherlands (HR per 25 nmol/L increment in predicted 25(OH)D, 0.20, 95% CI: 0.14-0.29) and the weakest observed for participants from Spain (HR per 25 nmol/L increment in predicted 25(OH)D, 0.68, 95% CI: 0.50-0.92) (Figure 15).

When analyses were limited to individuals who self-reported at baseline being in 'excellent' or 'good' health (n=65,390), similar associations were observed (sexes combined multivariable model: HR per 25 nmol/L increment in predicted 25(OH)D, 0.57, 95% CI: 0.46-0.72).

Table 30. Risk (hazard ratios) of all-cause mortality amongst men and women associated with predicted 25-hydroxyvitamin D level

	1	2	3	4	5		Per 25 nmol/L	<5 years excluded Per 25 nmol/L
			Quintile of predicted 25(OH)D (nmol/L)					
Men	<43.0	43.0-45.9	46.0-48.3	48.4-51.1	≥51.2		increase	increase
Women	<48.3	48.3-52.6	52.7-56.2	56.3-60.4	≥60.5 *P*-trend		increase	increase
Total all-cause mortality								
Both sexes								
N cases	4,335	3,546	2,949	2,651	2,780			
Person-years	721,056	762,567	787,833	798,436	780,817			
Basic model - HR (95% CI) †	1.00	0.81 (0.77-0.85)	0.71 (0.68-0.75)	0.66 (0.63-0.70)	0.60 (0.56-0.64)	<0.001	0.44 (0.41-0.48)	0.43 (0.39-0.47)
Multivariable model - HR (95% CI) ‡	1.00	0.87 (0.83-0.92)	0.78 (0.74-0.83)	0.72 (0.68-0.77)	0.65 (0.61-0.70)	<0.001	0.48 (0.43-0.53)	0.48 (0.42-0.53)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.92 (0.87-0.97)	0.86 (0.81-0.92)	0.82 (0.76-0.88)	0.75 (0.69-0.82)	<0.001	0.57 (0.50-0.64)	0.57 (0.49-0.66)
Men								
N cases	2,155	1,784	1,476	1,286	1,037			
Person-years	235,309	244,354	250,077	253,573	258,895			
Basic model - HR (95% CI) †	1.00	0.82 (0.77-0.88)	0.72 (0.67-0.78)	0.67 (0.62-0.72)	0.62 (0.57-0.67)	<0.001	0.35 (0.30-0.41)	0.35 (0.29-0.41)
Multivariable model - HR (95% CI) ‡	1.00	0.90 (0.84-0.96)	0.82 (0.76-0.89)	0.77 (0.71-0.84)	0.69 (0.63-0.76)	<0.001	0.46 (0.39-0.54)	0.48 (0.40-0.58)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.94 (0.88-1.02)	0.89 (0.81-0.98)	0.85 (0.76-0.94)	0.77 (0.68-0.87)	<0.001	0.49 (0.38-0.64)	0.53 (0.40-0.71)
Women								
N cases	2,180	1,762	1,473	1,365	1,743			
Person-years	485,747	518,213	537,756	544,863	521,922			
Basic model - HR (95% CI) †	1.00	0.80 (0.75-0.86)	0.70 (0.65-0.76)	0.65 (0.60-0.70)	0.58 (0.53-0.63)	<0.001	0.49 (0.45-0.54)	0.48 (0.43-0.54)
Multivariable model - HR (95% CI) ‡	1.00	0.83 (0.77-0.89)	0.73 (0.67-0.79)	0.67 (0.61-0.73)	0.59 (0.53-0.65)	<0.001	0.47 (0.41-0.53)	0.45 (0.39-0.52)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.88 (0.82-0.95)	0.81 (0.74-0.88)	0.76 (0.69-0.84)	0.70 (0.62-0.78)	<0.001	0.57 (0.49-0.66)	0.55 (0.47-0.65)

† Basic model - Cox regression stratified by age (1-year categories), sex, and centre.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

ɸ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Table 31. Risk (hazard ratios) of all-cause mortality associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level across strata of lifestyle, demographic, anthropometric, and dietary variables

	Total all-cause mortality	
	Both sexes	
Stratification variable		
	Predicted 25(OH)D HR (95% CI), per 25 nmol/L ‡	*P* interaction
Overall	0.48 (0.43-0.53)	
Sex		0.84
Men	0.46 (0.39-0.54)	
Women	0.47 (0.41-0.53)	
Age at recruitment		0.61
<55 years old	0.59 (0.49-0.70)	
55-<65 years old	0.46 (0.40-0.53)	
≥65 years old	0.39 (0.32-0.48)	
Follow-up		0.69
<5 years	0.48 (0.39-0.60)	
5-10 years	0.53 (0.46-0.62)	
≥10 years	0.49 (0.41-0.57)	
Smoking status		0.79
Never smoked	0.50 (0.42-0.59)	
Former smoker	0.46 (0.38-0.55)	
Current smoker	0.47 (0.40-0.55)	
Body mass index		0.001
<25.0 kg/m2	0.52 (0.45-0.61)	
25.0-29.9 kg/m2	0.49 (0.42-0.58)	
≥30 kg/m2	0.33 (0.26-0.41)	
Physical activity		0.11
Inactive	0.47 (0.38-0.60)	
Moderately inactive	0.59 (0.47-0.74)	
Moderately active	0.61 (0.46-0.82)	
Active	0.71 (0.53-0.95)	
Alcohol consumption		0.94
Non-consumers	0.50 (0.38-0.64)	
<15 g/day	0.45 (0.40-0.52)	
15-29.9 g/day	0.58 (0.43-0.77)	
≥30 g/day	0.45 (0.36-0.58)	
Red and processed meat consumption *		0.53
Below median	0.54 (0.47-0.63)	
Above median	0.43 (0.37-0.49)	
Fruit and vegetable consumption *		0.79
Below median	0.47 (0.41-0.54)	
Above median	0.49 (0.42-0.57)	
Fish consumption *		0.05
Below median	0.35 (0.29-0.41)	
Above median	0.55 (0.48-0.62)	
Calcium intake *		0.25
Below median	0.51 (0.44-0.59)	
Above median	0.45 (0.39-0.52)	
Fibre intake *		0.58
Below median	0.47 (0.41-0.54)	
Above median	0.49 (0.43-0.57)	
Retinol intake *		0.58
Below median	0.49 (0.42-0.57)	
Above median	0.47 (0.41-0.53)	
Menopausal status		0.47
Premenopausal	0.60 (0.41-0.90)	
Postmenopausal	0.44 (0.38-0.51)	
Perimenopausal	0.63 (0.41-0.99)	
Surgical postmenopausal	0.36 (0.20-0.64)	
Ever use of contraceptive pill		0.10
No	0.43 (0.37-0.51)	
Yes	0.51 (0.42-0.63)	

* Median intakes: red and processed meat=93.8 g/day in men and 59.1 g/day in women; fruit and vegetable=353 g/day in men and 436 g/day in women; fish consumption=26.6 g/day in men and 21.7 g/day in women; calcium=974 mg/day in men and 954 mg/day in women; fibre=23.9 g/day in men and 22.7 g/day in women; retinol=698 µg/day in men and 512 µg/day in women.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per

day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Table 32. Risk (hazard ratios) of all-cause mortality associated with the correlates/determinants of circulating 25(OH)D used in the predictor scores (mutually adjusted)

| | Total all-cause mortality | |
| | Both sexes | |
Predictor score	HR (95% CI)	P-trend
Dietary vitamin D (Per 2.5 μg/day)	1.00 (0.99-1.02)	
Physical activity		<0.001
Inactive	1.00	
Moderately inactive	0.84 (0.81-0.88)	
Moderately active	0.82 (0.78-0.86)	
Active	0.80 (0.76-0.84)	
Waist circumference (Per 1 cm)	1.02 (1.01-1.02)	
Country		<0.001
Denmark	1.00	
UK	0.84 (0.79-0.89)	
The Netherlands	0.84 (0.79-0.90)	
France	0.55 (0.50-0.61)	
Germany	0.90 (0.85-0.96)	
Greece	0.72 (0.66-0.79)	
Italy	0.60 (0.56-0.65)	
Spain	0.63 (0.58-0.68)	
Ever use of menopausal hormone therapy		0.41
No	1.00	
Yes	0.98 (0.93-1.03)	

Multivariable model - Cox regression with predictor score determinants mutually adjusted for each other, plus additional adjustment for body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Figure 15. Risk (hazard ratios) of all-cause mortality, by country, associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level

	Number of cases		HR (95% CI) per predicted 25 nmol/L increase
Denmark	3997		0.48 (0.40, 0.58)
France	618		0.44 (0.25, 0.78)
Germany	1732		0.63 (0.46, 0.87)
Greece	1223		0.52 (0.31, 0.85)
Italy	1279		0.44 (0.29, 0.66)
Spain	1414		0.68 (0.50, 0.92)
The Netherlands	1469		0.20 (0.14, 0.29)
UK - General population	4180		0.48 (0.40, 0.59)
UK - Health conscious	349		0.25 (0.13, 0.48)
ALL EPIC	16261		0.48 (0.43, 0.53)

Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

3.5.4 Predicted 25-hydroxyvitamin D levels and cause specific mortality risk

Cancer mortality

Elevated predicted circulating 25(OH)D was associated with a reduced cancer mortality risk in the sexes combined multivariable model (n=6,710 deaths: Q5 vs. Q1, HR, 0.77, 95% CI: 0.69-0.86; *P*-trend <0.001) (Table 33). In the continuous models, a 25 nmol/L higher predicted 25(OH)D level was associated with a 35% lower cancer mortality risk (95% CI: 0.56-0.75). When participants whose follow-up time was less than 5 years were excluded, similar risk estimates were observed. When the associations were assessed in men and women separately, identical significant 35% lower cancer mortality risks for both per increment in predicted 25(OH)D were observed (*P*-interaction 0.88) (Tables 34 and 35). When the overall cancer mortality analysis was limited to individuals who self-reported being in 'excellent' or 'good' health at baseline (n=65,390 participants), a slightly weaker non-significant inverse association was observed (HR per 25 nmol/L increment in predicted 25(OH)D, 0.76, 95% CI: 0.52-1.10).

Inverse predicted 25(OH)D-cancer mortality associations were observed across strata of all lifestyle, demographic, anthropometric, and dietary variables considered (Table 36). The interaction terms for predicted 25(OH)D with BMI and ever use of contraceptive pill were significant when inputted into the cancer mortality models. Inverse associations were observed in all countries, with risk estimates ranging from 0.31 to 0.87 per 25 nmol/L of predicted 25(OH)D (*P*-heterogeneity 0.016) (Figure 16).

Circulatory diseases mortality

At the end of the follow-up period, 3,641 circulatory disease deaths were recorded. In the multivariable models, a 41% reduced risk (95% CI: 0.50-0.68) of circulatory disease mortality was observed in the sexes combined models, when participants in the highest predicted 25(OH)D quintile were compared against those in the lowest quintile (*P*-trend <0.001) (Table 33). In the continuous models, a 61% lower circulatory disease mortality risk (95% CI: 0.31-0.48) was observed per 25 nmol/L increment in predicted 25(OH)D. The inverse associations were similar when participants whose follow-up time was less than 5 years were excluded. A slightly stronger inverse association was observed for women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.32, 95% CI: 0.24-0.43) compared to men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.42, 95% CI: 0.30-0.59); although, this difference was not significant (*P*-interaction 0.16) (Tables 34 and 35). When the circulatory disease mortality analysis was limited to individuals who self-reported being in 'excellent' or

'good' health at baseline (n=65,390 participants), a slightly weaker inverse relationship was observed (both sexes multivariable model: HR per 25 nmol/L increment in predicted 25(OH)D, 0.56, 95% CI: 0.34-0.94).

When circulatory disease deaths were further sub-categorised, similar associations in the sexes combined models were observed for ischaemic heart disease mortality (HR per 25 nmol/L increment in predicted 25(OH)D, 0.39, 95% CI: 0.27-0.55) and cerebrovascular disease mortality (HR per 25 nmol/L increment in predicted 25(OH)D, 0.35, 95% CI: 0.23-0.53) in the multivariable models.

Table 36 shows the circulatory mortality associations across strata of lifestyle, demographic, anthropometric, and dietary variables. Significant interactions were observed for predicted 25(OH)D with age at recruitment, smoking status, BMI, and alcohol consumption; however, the risk estimates for the across the strata of these variables were all within confidence intervals of each other and all of the associations were inverse. Inverse associations were observed in all countries, with risk estimates ranging from 0.13 to 0.79 per 25 nmol/L of predicted 25(OH)D (*P*-heterogeneity 0.08).

Respiratory diseases mortality

Over the follow-up period, 544 deaths caused by respiratory disease were recorded. Higher predicted 25(OH)D was associated with reduced risk of respiratory disease death (Q5 vs. Q1, HR 0.32, 95% CI: 0.22-0.48) in the sexes combined multivariable model (*P*-trend <0.001) (Table 33). In the equivalent continuous model, an 88% lower risk of respiratory disease mortality (95% CI: 0.07-0.21) was observed per 25 nmol/L increment in predicted 25(OH)D. Similar strong associations were observed when men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.13, 95% CI: 0.05-0.34) and women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.14, 95% CI: 0.07-0.28) were analysed separately (*P*-interaction 0.43) (Tables 34 and 35).

When the analysis was limited to individuals who self-reported being in 'excellent' or 'good' health at baseline (n=65,390 participants), a similar inverse relationship was observed (both sexes multivariable model: HR per 25 nmol/L increment in predicted 25(OH)D, 0.16, 95% CI: 0.03-0.90). When the multivariable model was not adjusted for BMI, the inverse association

weakened but remained significant (HR per 25 nmol/L increment in predicted 25(OH)D, 0.38, 95% CI: 0.25-0.59).

A significant interaction was observed for predicted 25(OH)D with smoking status (*P*-interaction 0.02); although, strong inverse associations were observed amongst never smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.26, 95% CI: 0.08-0.87) and current smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.13, 95% CI: 0.06-0.29). When analysing by country no heterogeneity in the relationship was detected (*P*-heterogeneity 0.87).

Digestive diseases mortality

At the end of the follow-up period, 511 digestive disease deaths had accrued. A 79% significantly reduced digestive disease mortality risk (95% CI: 0.14-0.32) was observed in the sexes combined multivariable model when individuals in the highest quintile of predicted 25(OH)D were compared versus those in the lowest (*P*-trend <0.001) (Table 33). In the continuous model, an 89% lower risk (95% CI: 0.06-0.20) was observed per 25 nmol/L increment in predicted 25(OH)D. When analysed by sex, the inverse association was slightly stronger for men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.08, 95% CI: 0.03-0.18), compared to women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.12, 95% CI: 0.06-0.26) (*P*-interaction 0.02) (Tables 34 and 35).

When the analysis was limited to individuals who self-reported being in 'excellent' or 'good' health at baseline (n=65,390 participants), a similar inverse relationship was observed (both sexes multivariable model: HR per 25 nmol/L increment in predicted 25(OH)D, 0.06, 95% CI: 0.02-0.27).

A significant interaction was observed for predicted 25(OH)D with smoking status (*P*-interaction 0.02); although, strong inverse associations were observed amongst never smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.05, 95% CI: 0.02-0.15) and current smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.16, 95% CI: 0.06-0.39). When analysed by country, similar associations were observed (*P*-heterogeneity 0.96).

External causes mortality

The associations between predicted 25(OH)D levels and deaths caused by external causes (n=313 deaths) (i.e. accidents and injuries) were investigated to act as "negative controls"; whereby, the exposure (predicted circulating 25(OH)D) and the end-points have no plausible mechanism for a causal effect. No association was observed in the sexes combined multivariable model (Q5 vs. Q1, HR 0.97, 95% CI: 0.71-1.32; *P*-trend 0.77) (Table 33). Null associations were also observed in the men (Q5 vs. Q1, HR 0.84, 95% CI: 0.57-1.23; *P*-trend 0.61) and women's (Q5 vs. Q1, HR 1.17, 95% CI: 0.68-2.01; *P*-trend 0.70) multivariable models (*P*-interaction 0.20) (Tables 34 and 35).

Table 33. Risk (hazard ratios) of cause-specific mortality amongst men and women associated with predicted 25-hydroxyvitamin D level

| | Quintile of predicted 25(OH)D (nmol/L) | | | | | | | <5 years excluded |
	1	2	3	4	5		Per 25 nmol/L increase	Per 25 nmol/L increase
Men	<43.0	43.0-45.9	46.0-48.3	48.4-51.1	≥51.2			
Women	<48.3	48.3-52.6	52.7-56.2	56.3-60.4	≥60.5	*P*-trend		
Person-years	698,807	726,295	744,828	752,787	726,586			
Cancer (C00-D48)								
N deaths	1,609	1,388	1,245	1,163	1,305			
Basic model - HR (95% CI) †	1.00	0.88 (0.82-0.95)	0.81 (0.75-0.88)	0.77 (0.71-0.84)	0.73 (0.66-0.80)	<0.001	0.62 (0.55-0.70)	0.65 (0.55-0.76)
Multivariable model - HR (95% CI) ‡	1.00	0.93 (0.86-1.00)	0.87 (0.79-0.95)	0.82 (0.75-0.91)	0.77 (0.69-0.86)	<0.001	0.65 (0.56-0.75)	0.70 (0.57-0.84)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.95 (0.88-1.04)	0.91 (0.82-1.00)	0.87 (0.78-0.98)	0.82 (0.72-0.94)	0.002	0.69 (0.57-0.83)	0.74 (0.59-0.95)
Circulatory diseases (I00-I99)								
N deaths	1,227	857	628	501	428			
Basic model - HR (95% CI) †	1.00	0.76 (0.69-0.84)	0.63 (0.57-0.70)	0.54 (0.48-0.61)	0.46 (0.40-0.52)	<0.001	0.26 (0.21-0.31)	0.26 (0.21-0.32)
Multivariable model - HR (95% CI) ‡	1.00	0.87 (0.79-0.96)	0.76 (0.68-0.86)	0.68 (0.60-0.77)	0.59 (0.50-0.68)	<0.001	0.39 (0.31-0.48)	0.39 (0.30-0.51)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.91 (0.82-1.02)	0.83 (0.73-0.94)	0.75 (0.64-0.87)	0.65 (0.54-0.79)	<0.001	0.43 (0.33-0.58)	0.44 (0.31-0.61)
Respiratory diseases (J30-J98)								
N deaths	166	131	69	65	113			
Basic model - HR (95% CI) †	1.00	0.81 (0.63-1.04)	0.44 (0.32-0.60)	0.41 (0.29-0.56)	0.45 (0.32-0.63)	<0.001	0.27 (0.18-0.41)	0.22 (0.13-0.35)
Multivariable model - HR (95% CI) ‡	1.00	0.78 (0.60-1.02)	0.42 (0.30-0.58)	0.36 (0.25-0.51)	0.32 (0.22-0.48)	<0.001	0.12 (0.07-0.21)	0.12 (0.06-0.22)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.94 (0.71-1.24)	0.59 (0.41-0.84)	0.56 (0.38-0.84)	0.60 (0.38-0.97)	0.007	0.28 (0.14-0.55)	0.23 (0.11-0.50)
Digestive diseases (K00-K93)								
N deaths	191	98	83	78	61			
Basic model - HR (95% CI) †	1.00	0.47 (0.36-0.60)	0.38 (0.29-0.50)	0.36 (0.27-0.48)	0.20 (0.14-0.29)	<0.001	0.11 (0.07-0.18)	0.13 (0.07-0.22)
Multivariable model - HR (95% CI) ‡	1.00	0.51 (0.39-0.66)	0.43 (0.32-0.57)	0.40 (0.29-0.55)	0.21 (0.14-0.32)	<0.001	0.11 (0.06-0.20)	0.17 (0.09-0.34)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.57 (0.43-0.76)	0.52 (0.37-0.73)	0.50 (0.34-0.75)	0.28 (0.17-0.47)	<0.001	0.19 (0.09-0.40)	0.27 (0.11-0.66)
External causes (S00-Y98)								
N deaths	174	132	159	122	155			
Basic model - HR (95% CI) †	1.00	0.81 (0.64-1.03)	1.07 (0.84-1.36)	0.83 (0.63-1.09)	1.04 (0.79-1.37)	0.85	1.00 (0.68-1.48)	0.71 (0.43-1.19)
Multivariable model - HR (95% CI) ‡	1.00	0.81 (0.63-1.04)	1.05 (0.81-1.36)	0.79 (0.59-1.06)	0.97 (0.71-1.32)	0.77	0.84 (0.53-1.33)	0.71 (0.39-1.29)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.79 (0.61-1.03)	1.02 (0.76-1.36)	0.76 (0.54-1.08)	0.94 (0.64-1.39)	0.73	0.75 (0.42-1.35)	0.65 (0.30-1.40)

† Basic model - Cox regression stratified by age (1-year categories), sex, and centre.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever

use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

ϕ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Table 34. Risk (hazard ratios) of cause-specific mortality amongst men associated with predicted 25-hydroxyvitamin D level

Men	Quantile of predicted 25(OH)D (nmol/L)					P-trend	Per 25 nmol/L increase
	1	2	3	4	5		
	<43.0	43.0-45.9	46.0-48.3	48.4-51.1	≥51.2		
Person-years	228,411	232,172	234,695	236,178	240,852		
Cancer (C00-D48)							
N deaths	776	636	583	515	413		
Basic model - HR (95% CI) †	1.00	0.86 (0.77-0.96)	0.84 (0.74-0.94)	0.80 (0.70-0.90)	0.73 (0.64-0.84)	<0.001	0.56 (0.44-0.71)
Multivariable model - HR (95% CI) ‡	1.00	0.92 (0.82-1.03)	0.91 (0.81-1.03)	0.87 (0.76-0.99)	0.79 (0.68-0.91)	0.002	0.65 (0.50-0.85)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.95 (0.84-1.07)	0.95 (0.82-1.11)	0.91 (0.77-1.09)	0.83 (0.68-1.02)	0.10	0.70 (0.47-1.04)
Circulatory diseases (I00-I99)							
N deaths	626	492	343	293	218		
Basic model - HR (95% CI) †	1.00	0.81 (0.71-0.91)	0.63 (0.54-0.73)	0.58 (0.50-0.68)	0.50 (0.42-0.59)	<0.001	0.21 (0.15-0.28)
Multivariable model - HR (95% CI) ‡	1.00	0.95 (0.84-1.08)	0.80 (0.69-0.93)	0.78 (0.66-0.92)	0.69 (0.57-0.84)	<0.001	0.42 (0.30-0.59)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.98 (0.85-1.13)	0.84 (0.70-1.01)	0.83 (0.67-1.02)	0.73 (0.57-0.95)	0.011	0.39 (0.23-0.65)
Respiratory diseases (J30-J98)							
N deaths	75	74	35	31	32		
Basic model - HR (95% CI) †	1.00	1.05 (0.75-1.47)	0.57 (0.37-0.88)	0.57 (0.36-0.90)	0.74 (0.47-1.17)	0.018	0.29 (0.13-0.66)
Multivariable model - HR (95% CI) ‡	1.00	0.97 (0.68-1.38)	0.50 (0.32-0.80)	0.48 (0.29-0.79)	0.51 (0.31-0.86)	0.001	0.13 (0.05-0.34)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.11 (0.75-1.63)	0.64 (0.37-1.10)	0.69 (0.37-1.28)	0.87 (0.42-1.78)	0.37	0.23 (0.05-1.03)
Digestive diseases (K00-K93)							
N deaths	117	58	48	39	21		
Basic model - HR (95% CI) †	1.00	0.48 (0.34-0.66)	0.39 (0.27-0.55)	0.32 (0.22-0.48)	0.19 (0.12-0.31)	<0.001	0.04 (0.02-0.10)
Multivariable model - HR (95% CI) ‡	1.00	0.55 (0.39-0.78)	0.48 (0.33-0.71)	0.41 (0.26-0.62)	0.22 (0.13-0.37)	<0.001	0.08 (0.03-0.18)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.60 (0.41-0.87)	0.55 (0.35-0.87)	0.45 (0.26-0.79)	0.23 (0.11-0.47)	<0.001	0.08 (0.02-0.34)
External causes (S00-Y98)							
N deaths	107	70	99	77	76		
Basic model - HR (95% CI) †	1.00	0.70 (0.51-0.96)	1.06 (0.79-1.44)	0.85 (0.61-1.19)	0.89 (0.63-1.26)	0.79	0.89 (0.48-1.62)
Multivariable model - HR (95% CI) ‡	1.00	0.70 (0.51-0.96)	1.07 (0.78-1.47)	0.83 (0.58-1.19)	0.84 (0.57-1.23)	0.61	0.79 (0.40-1.54)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.68 (0.48-0.97)	1.04 (0.70-1.52)	0.81 (0.51-1.27)	0.84 (0.49-1.42)	0.73	0.73 (0.27-1.97)

† Basic model - Cox regression stratified by age (1-year categories) and centre.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown), and

intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), and centre.

ϕ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Table 35. Risk (hazard ratios) of cause-specific mortality amongst women associated with predicted 25-hydroxyvitamin D level

Women	1	2	3	4	5	P-trend	Per 25 nmol/L increase
	<48.3	48.3-52.6	52.7-56.2	56.3-60.4	≥60.5		
Person-years	470,397	494,123	510,133	516,610	485,734		
Cancer (C00-D48)							
N deaths	833	752	662	648	892		
Basic model - HR (95% CI) †	1.00	0.89 (0.80-0.99)	0.79 (0.70-0.88)	0.75 (0.66-0.85)	0.72 (0.63-0.82)	<0.001	0.65 (0.56-0.75)
Multivariable model - HR (95% CI) ‡	1.00	0.92 (0.82-1.03)	0.82 (0.72-0.93)	0.78 (0.68-0.89)	0.74 (0.63-0.87)	<0.001	0.65 (0.54-0.78)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.95 (0.85-1.07)	0.87 (0.76-1.00)	0.83 (0.72-0.97)	0.81 (0.67-0.97)	0.01	0.70 (0.57-0.87)
Circulatory diseases (I00-I99)							
N deaths	601	365	285	208	210		
Basic model - HR (95% CI) †	1.00	0.71 (0.61-0.82)	0.63 (0.53-0.74)	0.49 (0.40-0.59)	0.39 (0.31-0.49)	<0.001	0.29 (0.23-0.37)
Multivariable model - HR (95% CI) ‡	1.00	0.77 (0.66-0.90)	0.71 (0.59-0.84)	0.55 (0.44-0.67)	0.45 (0.35-0.57)	<0.001	0.32 (0.24-0.43)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.81 (0.69-0.95)	0.78 (0.64-0.95)	0.62 (0.50-0.78)	0.53 (0.40-0.70)	<0.001	0.38 (0.27-0.55)
Respiratory diseases (J30-J98)							
N deaths	91	57	34	34	81		
Basic model - HR (95% CI) †	1.00	0.58 (0.40-0.83)	0.31 (0.20-0.49)	0.27 (0.17-0.42)	0.27 (0.17-0.42)	<0.001	0.26 (0.16-0.43)
Multivariable model - HR (95% CI) ‡	1.00	0.55 (0.38-0.82)	0.30 (0.19-0.48)	0.22 (0.13-0.37)	0.18 (0.10-0.32)	<0.001	0.14 (0.07-0.28)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.66 (0.44-0.98)	0.42 (0.26-0.70)	0.35 (0.20-0.62)	0.36 (0.19-0.68)	0.001	0.39 (0.18-0.85)
Digestive diseases (K00-K93)							
N deaths	74	40	35	39	40		
Basic model - HR (95% CI) †	1.00	0.45 (0.30-0.68)	0.38 (0.24-0.59)	0.41 (0.26-0.64)	0.22 (0.13-0.38)	<0.001	0.19 (0.11-0.35)
Multivariable model - HR (95% CI) ‡	1.00	0.44 (0.28-0.67)	0.35 (0.22-0.57)	0.37 (0.22-0.62)	0.18 (0.09-0.35)	<0.001	0.12 (0.06-0.26)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.49 (0.31-0.77)	0.44 (0.26-0.74)	0.50 (0.28-0.89)	0.27 (0.13-0.58)	0.002	0.19 (0.07-0.48)
External causes (S00-Y98)							
N deaths	67	62	60	45	79		
Basic model - HR (95% CI) †	1.00	1.01 (0.69-1.48)	1.09 (0.73-1.65)	0.81 (0.51-1.29)	1.34 (0.84-2.13)	0.33	1.09 (0.66-1.82)
Multivariable model - HR (95% CI) ‡	1.00	0.97 (0.65-1.45)	1.03 (0.66-1.61)	0.74 (0.44-1.23)	1.17 (0.68-2.01)	0.70	0.87 (0.46-1.64)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.95 (0.62-1.43)	0.99 (0.61-1.60)	0.70 (0.40-1.24)	1.11 (0.60-2.05)	0.79	0.77 (0.36-1.63)

† Basic model - Cox regression stratified by age (1-year categories) and centre.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m²), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day;

former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), and centre.

ɸ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Table 36. Risk (hazard ratios) of cancer and circulatory disease mortality associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level across strata of lifestyle, demographic, anthropometric, and dietary variables
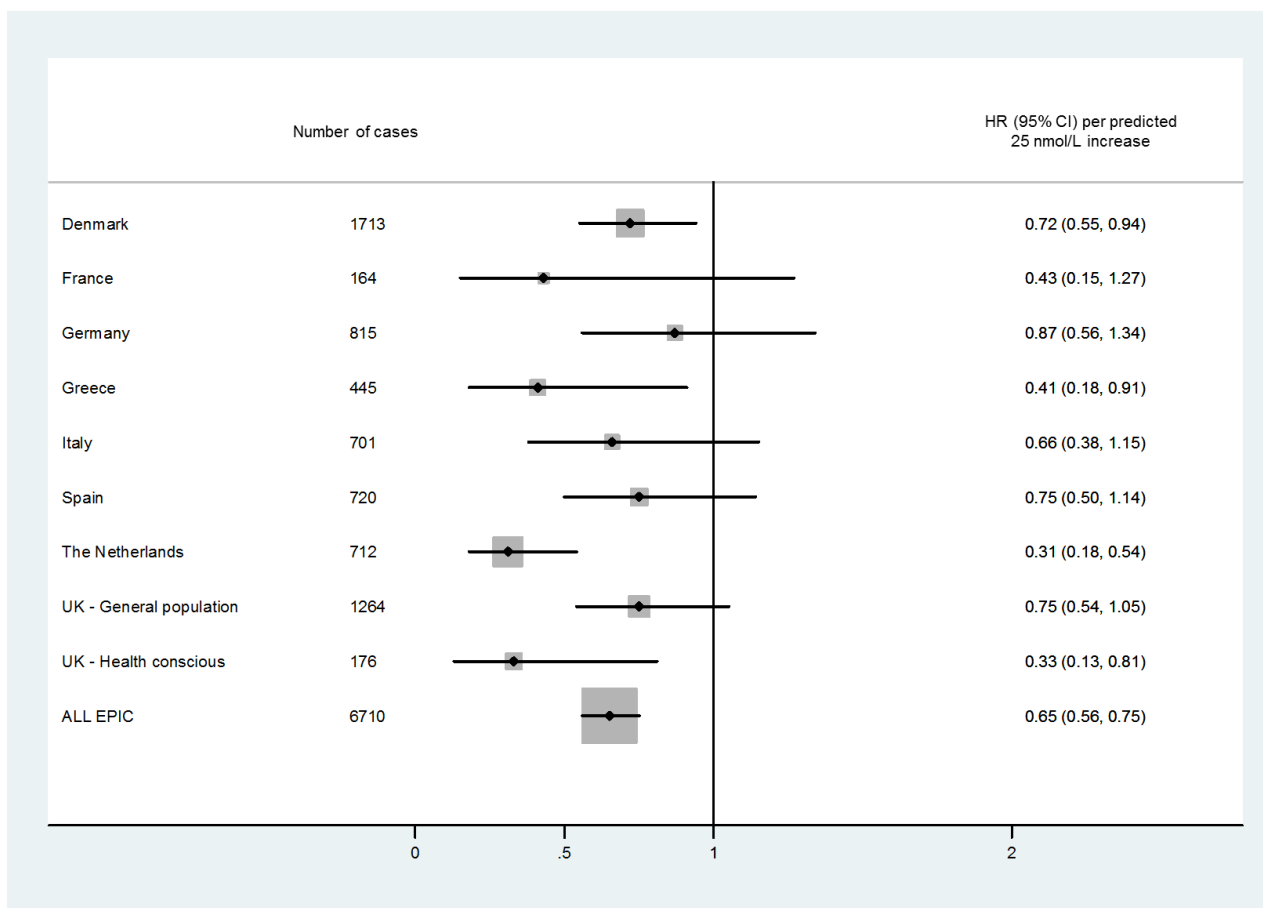
Stratification variable	Cancer mortality Both sexes		Circulatory disease mortality Both sexes	
	Predicted 25(OH)D HR (95% CI), per 25 nmol/L ‡	P interaction	Predicted 25(OH)D HR (95% CI), per 25 nmol/L ‡	P interaction
Overall	0.65 (0.56-0.75)		0.39 (0.31-0.48)	
Sex		0.88		0.16
Men	0.65 (0.50-0.85)		0.42 (0.30-0.59)	
Women	0.65 (0.54-0.78)		0.32 (0.24-0.43)	
Age at recruitment		0.62		0.04
<55 years old	0.73 (0.57-0.94)		0.45 (0.28-0.72)	
55-<65 years old	0.63 (0.51-0.77)		0.33 (0.24-0.45)	
≥65 years old	0.58 (0.39-0.87)		0.45 (0.31-0.65)	
Follow-up		0.80		0.82
<5 years	0.62 (0.45-0.84)		0.38 (0.25-0.58)	
5-10 years	0.69 (0.56-0.86)		0.45 (0.33-0.62)	
≥10 years	0.71 (0.53-0.94)		0.42 (0.27-0.65)	
Smoking status		0.92		0.001
Never smoked	0.75 (0.58-0.96)		0.29 (0.20-0.42)	
Former smoker	0.60 (0.44-0.80)		0.41 (0.27-0.61)	
Current smoker	0.61 (0.47-0.78)		0.46 (0.32-0.67)	
Body mass index		0.02		0.02
<25.0 kg/m2	0.77 (0.61-0.96)		0.33 (0.23-0.47)	
25.0-29.9 kg/m2	0.71 (0.56-0.91)		0.42 (0.30-0.60)	
≥30 kg/m2	0.36 (0.25-0.52)		0.26 (0.17-0.41)	
Physical activity		0.36		0.03
Inactive	0.51 (0.35-0.75)		0.39 (0.24-0.64)	
Moderately inactive	0.73 (0.52-1.02)		0.67 (0.40-1.13)	
Moderately active	0.79 (0.51-1.20)		0.28 (0.14-0.56)	
Active	0.80 (0.53-1.20)		0.41 (0.20-0.87)	
Alcohol consumption		0.45		0.02
Non-consumers	0.73 (0.49-1.09)		0.17 (0.09-0.32)	
<15 g/day	0.60 (0.49-0.73)		0.42 (0.32-0.56)	
15-29.9 g/day	0.70 (0.46-1.08)		0.41 (0.22-0.77)	
≥30 g/day	0.70 (0.49-0.99)		0.39 (0.22-0.70)	
Red and processed meat consumption *		0.51		0.52
Below median	0.78 (0.63-0.97)		0.37 (0.27-0.51)	
Above median	0.54 (0.44-0.67)		0.42 (0.31-0.57)	
Fruit and vegetable consumption *		0.75		0.57
Below median	0.70 (0.57-0.85)		0.39 (0.29-0.53)	
Above median	0.59 (0.47-0.75)		0.40 (0.29-0.55)	
Fish consumption *		0.87		0.35
Below median	0.52 (0.40-0.68)		0.29 (0.21-0.42)	
Above median	0.71 (0.59-0.87)		0.43 (0.33-0.58)	
Calcium intake *		0.93		0.60
Below median	0.66 (0.54-0.82)		0.41 (0.30-0.56)	
Above median	0.63 (0.51-0.78)		0.37 (0.27-0.50)	
Fibre intake *		0.67		0.68
Below median	0.77 (0.63-0.95)		0.36 (0.27-0.49)	
Above median	0.54 (0.44-0.68)		0.42 (0.31-0.58)	
Retinol intake *		0.96		0.09
Below median	0.71 (0.56-0.89)		0.33 (0.24-0.46)	
Above median	0.59 (0.49-0.73)		0.43 (0.32-0.58)	
Menopausal status		0.91		0.12
Premenopausal	0.68 (0.41-1.13)		0.31 (0.08-1.11)	
Postmenopausal	0.62 (0.49-0.77)		0.30 (0.22-0.42)	
Perimenopausal	0.80 (0.43-1.46)		1.04 (0.29-3.79)	
Surgical postmenopausal	0.64 (0.27-1.54)		0.56 (0.14-2.23)	
Ever use of contraceptive pill		0.02		0.30
No	0.50 (0.39-0.65)		0.36 (0.25-0.51)	
Yes	0.81 (0.62-1.06)		0.25 (0.15-0.43)	

* Median intakes: red and processed meat=93.8 g/day in men and 59.1 g/day in women; fruit and vegetable=353 g/day in men and 436 g/day in women; fish consumption=26.6 g/day in men and 21.7 g/day in women; calcium=974 mg/day in men and 954 mg/day in women; fibre=23.9 g/day in men and 22.7 g/day in women; retinol=698 µg/day in men and 512 µg/day in women.

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or

surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

Figure 16. Risk (hazard ratios) of cancer mortality, by country, associated with a 25 nmol/L increment of predicted 25-hydroxyvitamin D level

	Number of cases		HR (95% CI) per predicted 25 nmol/L increase
Denmark	1713		0.72 (0.55, 0.94)
France	164		0.43 (0.15, 1.27)
Germany	815		0.87 (0.56, 1.34)
Greece	445		0.41 (0.18, 0.91)
Italy	701		0.66 (0.38, 1.15)
Spain	720		0.75 (0.50, 1.14)
The Netherlands	712		0.31 (0.18, 0.54)
UK - General population	1264		0.75 (0.54, 1.05)
UK - Health conscious	176		0.33 (0.13, 0.81)
ALL EPIC	6710		0.65 (0.56, 0.75)

Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

3.5.5 Predicted 25-hydroxyvitamin D levels and smoking related and non-smoking related cancer mortality

Smoking related cancer mortality

Elevated predicted circulating 25(OH)D was associated with a reduced risk of smoking related cancer deaths (n=3,681 deaths: Q5 vs. Q1, HR 0.73, 95% CI: 0.63-0.84) in the sexes combined multivariable model (*P*-trend <0.001) (Table 37). In the continuous models, a 38% lower risk (95% CI: 0.51-0.77) of smoking related cancer mortality was observed, and similar associations were also observed when participants whose follow-up time was less than 5 years were excluded from the analysis. A stronger inverse association was observed amongst women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.55, 95% CI: 0.41-0.72) than for men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.73, 95% CI: 0.53-1.00); although this difference was non-significant (*P*-interaction 0.21).

When the association was analysed by smoking status, similar inverse associations were observed amongst never smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.67, 95% CI: 0.44-1.00); former smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.67, 95% CI: 0.45-0.99) and current smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.60, 95% CI: 0.45-0.82) (*P*-interaction 0.88). No heterogeneity between countries was found (*P*-heterogeneity 0.20), with risk estimates below 1 observed in all countries (HRs per 25 nmol/L increment in predicted 25(OH)D ranging from HR 0.23 to 0.96).

Non-smoking related cancer mortality

A 16% reduced risk (95% CI: 0.71-0.99) of non-smoking related cancer mortality (n=3,029 deaths) was observed in the sexes combined multivariable model (*P*-trend 0.017) (Table 37). In the equivalent continuous model, a 32% lower risk (95% CI: 0.55-0.85) was observed per 25 nmol/L increment in predicted circulating 25(OH)D. Inverse associations of similar magnitude were observed when participants whose follow-up times were less than 5 years were excluded. The inverse association was stronger for men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.51, 95% CI: 0.32-0.82) than for women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.74, 95% CI: 0.58-0.95), although this difference was non-significant (*P*-interaction 0.15). Inverse associations of similar magnitude were observed across all countries (*P*-heterogeneity 0.36).

Table 37. Risk (hazard ratios) of smoking related and non-smoking related cancer mortality associated with predicted 25-hydroxyvitamin D level

	Quintile of predicted 25(OH)D (nmol/L)							<5 years excluded
	1	2	3	4	5		Per 25 nmol/L	Per 25 nmol/L
Men	<43.0	43.0-45.9	46.0-48.3	48.4-51.1	≥51.2	*P*-trend	increase	increase
Women	<48.3	48.3-52.6	52.7-56.2	56.3-60.4	≥60.5			
Smoking related cancers *								
Both sexes								
N deaths	919	757	687	611	707			
Multivariable model - HR (95% CI) ‡	1.00	0.91 (0.82-1.02)	0.87 (0.78-0.98)	0.80 (0.70-0.90)	0.73 (0.63-0.84)	<0.001	0.62 (0.51-0.77)	0.69 (0.53-0.89)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.94 (0.84-1.05)	0.92 (0.80-1.05)	0.83 (0.71-0.97)	0.76 (0.63-0.91)	0.002	0.67 (0.51-0.87)	0.78 (0.56-1.07)
Men								
N deaths	541	437	401	337	271			
Multivariable model - HR (95% CI) ‡	1.00	0.96 (0.84-1.10)	0.96 (0.82-1.11)	0.88 (0.75-1.03)	0.82 (0.68-0.98)	0.02	0.73 (0.53-1.00)	
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.98 (0.84-1.14)	0.99 (0.82-1.19)	0.90 (0.73-1.11)	0.81 (0.63-1.04)	0.10	0.76 (0.47-1.22)	
Women								
N deaths	378	320	286	274	436			
Multivariable model - HR (95% CI) ‡	1.00	0.83 (0.70-0.98)	0.74 (0.61-0.90)	0.66 (0.53-0.81)	0.58 (0.46-0.74)	<0.001	0.55 (0.41-0.72)	
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.86 (0.72-1.02)	0.79 (0.64-0.97)	0.71 (0.56-0.89)	0.64 (0.48-0.84)	0.001	0.61 (0.44-0.84)	
Non-smoking related cancers *								
Both sexes								
N deaths	690	631	558	552	598			
Multivariable model - HR (95% CI) ‡	1.00	0.95 (0.84-1.07)	0.86 (0.75-0.98)	0.87 (0.75-1.00)	0.84 (0.71-0.99)	0.017	0.68 (0.55-0.85)	0.72 (0.54-0.97)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.98 (0.86-1.11)	0.90 (0.78-1.05)	0.92 (0.78-1.09)	0.90 (0.74-1.10)	0.26	0.72 (0.55-0.93)	0.71 (0.49-1.02)
Men								
N deaths	235	199	182	178	142			
Multivariable model - HR (95% CI) ‡	1.00	0.85 (0.69-1.04)	0.82 (0.66-1.02)	0.84 (0.66-1.06)	0.73 (0.57-0.95)	0.03	0.51 (0.32-0.82)	
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.87 (0.70-1.08)	0.85 (0.65-1.12)	0.91 (0.67-1.23)	0.84 (0.59-1.20)	0.47	0.55 (0.27-1.13)	
Women								
N deaths	455	432	376	374	456			
Multivariable model - HR (95% CI) ‡	1.00	1.01 (0.87-1.17)	0.88 (0.75-1.05)	0.88 (0.73-1.06)	0.89 (0.72-1.11)	0.18	0.74 (0.58-0.95)	
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	1.04 (0.89-1.22)	0.94 (0.78-1.13)	0.95 (0.77-1.17)	0.97 (0.76-1.24)	0.65	0.79 (0.59-1.06)	

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

ɸ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Smoking related cancers were: oral cavity (C01-C06, C08); oropharynx (C09, C10, C12-C14); nasopharynx (C11); oesophagus (C15); stomach (C16); colorectal (C18-C20); liver (C22); pancreas (C25); nasal cavity and sinuses (C300, C31); larynx (C32); lung (C34); kidney (C64); bladder (C65, C67); and myeloid leukemia (C92). Non-smoking related cancers were all other cancers.

3.5.6 Predicted 25-hydroxyvitamin D levels and digestive system and non-digestive system cancer mortality

Digestive system cancer mortality

Higher predicted circulating 25(OH)D was associated with a non-significant reduced digestive system cancer mortality risk in the sexes combined multivariable model (n=1,629 deaths: Q5 vs. Q1, HR 0.81, 95% CI: 0.65-1.01), with a near significance linear trend (*P*-trend 0.06) (Table 38). In the equivalent continuous model, a 25 nmol/L higher predicted 25(OH)D level was associated with a 33% lower digestive system cancer mortality risk (95% CI: 0.49-0.91). When participants whose follow-up time was less than 5 years were excluded, similar risk estimates were yielded. This association was stronger and significant for women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.57, 95% CI: 0.38-0.84) compared to men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.79, 95% CI: 0.47-1.29); although this difference was not significant (*P*-interaction 0.25). No heterogeneity was observed by country for the inverse association observed (*P*-heterogeneity 0.84).

Non-digestive cancer mortality

In the sexes combined multivariable model, a 24% reduced (95% CI: 0.67-0.86) non-digestive system mortality risk (n=5,074 deaths) was observed when participants in the highest and lowest predicted circulating 25(OH)D quintiles were compared (Table 38). In the equivalent continuous models, a 36% lower risk (95% CI: 0.54-0.76) was also observed per 25 nmol/L increment in predicted 25(OH)D. Inverse associations of similar magnitude were observed when participants whose follow-up time was less than 5 years were excluded. Similar associations were observed when men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.60, 95% CI: 0.44-0.82) and women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.67, 95% CI: 0.54-0.83) were analysed separately (*P*-interaction 0.57). Significant heterogeneity was observed in the inverse associations when analysed by country (*P*-heterogeneity 0.007); although HRs below 1 per 25 nmol/L increment in predicted 25(OH)D were observed in all countries (ranging from HR 0.25 to 0.95).

Table 38. Risk (hazard ratios) of digestive and non-digestive cancer mortality associated with predicted 25-hydroxyvitamin D level

	Quintile of predicted 25(OH)D (nmol/L)							<5 years excluded
	1	2	3	4	5		Per 25 nmol/L	Per 25 nmol/L
Men	<43.0	43.0-45.9	46.0-48.3	48.4-51.1	≥51.2	*P*-trend	increase	increase
Women	<48.3	48.3-52.6	52.7-56.2	56.3-60.4	≥60.5			
Digestive system cancers *								
Both sexes								
N deaths	410	354	293	284	288			
Multivariable model - HR (95% CI) ‡	1.00	0.97 (0.83-1.14)	0.88 (0.74-1.05)	0.90 (0.75-1.09)	0.81 (0.65-1.01)	0.06	0.67 (0.49-0.91)	0.68 (0.47-0.99)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.02 (0.86-1.20)	0.95 (0.77-1.16)	0.97 (0.78-1.22)	0.86 (0.66-1.13)	0.32	0.70 (0.48-1.03)	0.73 (0.46-1.17)
Men								
N deaths	202	198	136	135	120			
Multivariable model - HR (95% CI) ‡	1.00	1.13 (0.91-1.39)	0.85 (0.66-1.08)	0.91 (0.70-1.18)	0.92 (0.70-1.23)	0.28	0.79 (0.47-1.29)	
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	1.08 (0.86-1.37)	0.79 (0.59-1.06)	0.82 (0.58-1.15)	0.78 (0.52-1.15)	0.10	0.56 (0.26-1.20)	
Women								
N deaths	208	156	157	149	168			
Multivariable model - HR (95% CI) ‡	1.00	0.81 (0.64-1.03)	0.89 (0.69-1.16)	0.87 (0.65-1.16)	0.68 (0.49-0.96)	0.07	0.57 (0.38-0.84)	
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.89 (0.70-1.14)	1.04 (0.78-1.38)	1.06 (0.77-1.45)	0.86 (0.58-1.27)	0.65	0.70 (0.44-1.12)	
Non-digestive system cancers *								
Both sexes								
N deaths	1199	1034	952	872	1017			
Multivariable model - HR (95% CI) ‡	1.00	0.91 (0.83-0.99)	0.86 (0.78-0.95)	0.80 (0.72-0.89)	0.76 (0.67-0.86)	<0.001	0.64 (0.54-0.76)	0.70 (0.56-0.88)
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.93 (0.84-1.03)	0.90 (0.80-1.01)	0.84 (0.74-0.95)	0.80 (0.69-0.94)	0.003	0.68 (0.55-0.84)	0.74 (0.56-0.98)
Men								
N deaths	574	438	447	380	293			
Multivariable model - HR (95% CI) ‡	1.00	0.85 (0.74-0.97)	0.93 (0.80-1.07)	0.85 (0.73-0.99)	0.74 (0.62-0.88)	0.003	0.60 (0.44-0.82)	
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.89 (0.77-1.03)	1.01 (0.85-1.20)	0.95 (0.78-1.16)	0.85 (0.67-1.09)	0.35	0.76 (0.48-1.21)	
Women								
N deaths	625	596	505	499	724			
Multivariable model - HR (95% CI) ‡	1.00	0.96 (0.84-1.09)	0.80 (0.69-0.93)	0.75 (0.64-0.88)	0.76 (0.63-0.91)	<0.001	0.67 (0.54-0.83)	
Multivariable model + physical activity adj - HR (95% CI) φ	1.00	0.97 (0.85-1.11)	0.82 (0.70-0.96)	0.78 (0.65-0.93)	0.79 (0.64-0.97)	0.01	0.70 (0.55-0.90)	

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

171

ϕ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

Digestive system cancers were: oesophagus (C15); stomach (C16); colorectal (C18-C20); anus and anal canal (C21); liver and intrahepatic bile acids (C22); gallbladder (C23); other and unspecified parts of biliary tract (C24); pancreas (C25); and other ill-defined digestive organs (C26). Non-digestive system cancers were all other cancers.

3.5.7 Predicted 25-hydroxyvitamin D levels and mortality from individual cancers

Colorectal cancer mortality

By the end of the follow-up period, 491 colorectal cancer deaths had been recorded. In the sexes combined multivariable model, a non-significant 20% reduced colorectal cancer mortality risk (95% CI: 0.53-1.21) was observed amongst participants in the highest quintile when compared against those in the lowest quintile of predicted circulating 25(OH)D (*P*-trend 0.35) (Table 39). A non-significant 24% lower risk per 25 nmol/L increment in predicted circulating 25(OH)D was observed in the continuous model. When analysed separately, similar non-significant inverse associations were observed for men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.78, 95% CI: 0.28-2.19) and women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.67, 95% CI: 0.34-1.30) (*P*-interaction 0.84).

Pancreatic cancer mortality

Higher predicted 25(OH)D levels were associated with a 35% reduced risk (95% CI: 0.44-0.95) of pancreatic cancer mortality (n=540 deaths) (*P*-trend 0.036) (Table 39). Significant inverse associations were also observed in the continuous models; with similar associations observed when participants whose follow-up times were less than 5 years were excluded. When analysed by sex, non-significant inverse associations of similar strength were observed for men (n=248 deaths; HR per 25 nmol/L increment in predicted 25(OH)D, 0.66, 95% CI: 0.27-1.61) and women (n=292 deaths; HR per 25 nmol/L increment in predicted 25(OH)D, 0.52, 95% CI: 0.27-1.02) (*P*-interaction 0.72).

Lung cancer mortality

Over the follow-up period, 1,480 lung cancer deaths were recorded. In the sexes combined multivariable model, a 30% reduced lung cancer mortality risk (95% CI: 0.56-0.88) was observed when participants in the highest and lowest quintiles of predicted circulating 25(OH)D were compared (*P*-trend 0.002) (Table 39). In the equivalent continuous model, a non-significant 24% lower risk (95% CI: 0.55-1.05) was observed per 25 nmol/L increment in predicted 25(OH)D levels. Similar associations were observed when participants whose follow-up times were less than 5 years were excluded. Stronger inverse associations were observed amongst women (HR per 25 nmol/L increment in predicted 25(OH)D, 0.69, 95% CI: 0.45-1.06) than in men (HR per 25 nmol/L increment in predicted 25(OH)D, 0.83, 95% CI: 0.51-1.34), although this difference was non-significant (*P*-interaction 0.59).

However, the inverse lung cancer incidence association was absent when the multivariable model was not adjusted for BMI (both sexes: HR per 25 nmol/L increment in predicted 25(OH)D, 1.22, 95% CI: 0.94-1.58) (Table A2 – appendix).

When the association was analysed by smoking status, non-significant inverse associations were observed amongst never smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.76, 95% CI: 0.27-2.14) and current smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 0.72, 95% CI: 0.49-1.08); whilst a non-significant positive association was observed amongst former smokers (HR per 25 nmol/L increment in predicted 25(OH)D, 1.10, 95% CI: 0.57-2.11). Overall the association across strata of smoking status was non-significant (*P*-interaction 0.67).

Prostate cancer mortality
A 50% reduced (95% CI: 0.30-0.85) prostate cancer mortality risk (n=228 deaths) was observed in the categorical multivariable model (*P*-trend 0.02) (Table 39). In the equivalent continuous model, a 68% lower risk (95% CI: 0.12-0.84) was observed per 25 nmol/L increment in predicted circulating 25(OH)D. A stronger inverse association was observed when men with follow-up time less than 5 years were excluded (HR per 25 nmol/L increment in predicted 25(OH)D, 0.31, 95% CI: 0.09-1.09).

Breast cancer mortality
In the multivariable model, when the highest and lowest quintiles of predicted circulating 25(OH)D were compared, a non-significant 15% reduced risk (95% CI: 0.57-1.29) of breast cancer mortality (n=575 deaths) was observed (*P*-trend 0.44) (Table 39). In the continuous models, a non-significant 26% lower risk was observed per 25 nmol/L increment in predicted circulating 25(OH)D.

Table 39. Risk (hazard ratios) for mortality of specific cancer types amongst men and women by predicted 25-hydroxyvitamin D (25(OH)D) categories

Both sexes	Quintile of predicted 25(OH)D (nmol/L)							<5 years excluded
	1	2	3	4	5			
Men	<43.0	43.0-45.9	46.0-48.3	48.4-51.1	≥51.2		Per 25 nmol/L	Per 25 nmol/L
Women	<48.3	48.3-52.6	52.7-56.2	56.3-60.4	≥60.5	*P*-trend	increase	increase
Colorectal cancer (C18-C20)								
N deaths	120	103	96	86	86			
Multivariable model - HR (95% CI) ‡ ¥	1.00	0.95 (0.70-1.27)	0.97 (0.70-1.34)	0.92 (0.64-1.32)	0.80 (0.53-1.21)	0.35	0.76 (0.44-1.32)	0.84 (0.44-1.59)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	1.03 (0.75-1.40)	1.12 (0.78-1.61)	1.09 (0.72-1.66)	0.97 (0.59-1.58)	0.97	1.00 (0.52-1.94)	1.33 (0.63-2.83)
Pancreatic cancer (C25)								
N deaths	142	112	87	103	96			
Multivariable model - HR (95% CI) ‡	1.00	0.85 (0.64-1.12)	0.68 (0.50-0.94)	0.84 (0.60-1.16)	0.65 (0.44-0.95)	0.036	0.58 (0.34-0.99)	0.50 (0.26-0.95)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.89 (0.66-1.19)	0.73 (0.51-1.04)	0.89 (0.60-1.31)	0.68 (0.43-1.08)	0.15	0.65 (0.33-1.25)	0.53 (0.23-1.18)
Lung cancer (C34)								
N deaths	335	267	295	243	340			
Multivariable model - HR (95% CI) ‡	1.00	0.85 (0.71-1.01)	0.94 (0.78-1.14)	0.75 (0.61-0.92)	0.70 (0.56-0.88)	0.002	0.76 (0.55-1.05)	0.81 (0.54-1.21)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.87 (0.72-1.04)	0.97 (0.79-1.20)	0.78 (0.61-0.99)	0.73 (0.55-0.97)	0.03	0.91 (0.61-1.36)	1.02 (0.62-1.67)
Prostate cancer (C61)								
N deaths	61	49	41	47	30			
Multivariable model - HR (95% CI) ‡	1.00	0.71 (0.47-1.07)	0.60 (0.39-0.95)	0.70 (0.45-1.10)	0.50 (0.30-0.85)	0.02	0.32 (0.12-0.84)	0.31 (0.09-1.09)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.70 (0.45-1.10)	0.59 (0.34-1.02)	0.67 (0.36-1.24)	0.48 (0.23-1.00)	0.08	0.25 (0.05-1.18)	0.09 (0.01-0.73)
Breast cancer (C50)								
N deaths	120	106	110	98	141			
Multivariable model - HR (95% CI) ‡ ¶	1.00	0.88 (0.65-1.18)	0.90 (0.65-1.25)	0.78 (0.54-1.12)	0.85 (0.57-1.29)	0.44	0.74 (0.46-1.19)	0.82 (0.37-1.79)
Multivariable model + physical activity adj - HR (95% CI) ɸ	1.00	0.96 (0.71-1.31)	1.05 (0.74-1.50)	0.94 (0.63-1.41)	1.07 (0.67-1.70)	0.78	0.95 (0.55-1.63)	0.91 (0.36-2.27)

‡ Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

¥ Multivariable model – plus adjustment for cereal fibre intake (g/day).

¶ Multivariable model – plus adjustment for age at first pregnancy (<21; 21-<30; 30+ years old; no children; or not specified); and age at menarche (<12; 12-<15; 15+ years old; or not specified).

ɸ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

4. DISCUSSION

4.1 Dietary vitamin D intake and all-cause mortality, cancer mortality and cause-specific mortality

This analysis investigated the relationships between dietary vitamin D intake and all-cause mortality, cancer mortality and cause-specific mortality risk amongst participants in the EPIC cohort.

No associations were observed for dietary vitamin D intake and all of mortality end-points considered. The methodological limitations of using dietary intakes as a surrogate marker of vitamin D status are probably the cause of these null results. Firstly, as the EPIC derived vitamin D predictor scores suggest, intake of vitamin D from food sources contributes little to serum 25(OH)D status (1% of variance in circulating 25(OH)D for men and negligible variance for women). Secondly, recorded dietary vitamin D intakes within EPIC are low. For instance, men and women in the highest intake quintile had vitamin D intakes over 6.5 µg/day and 5.1 µg/day respectively. Such daily intakes would have little impact on endogenous vitamin D levels, with 1 µg of vitamin D intake being shown to increase circulating 25(OH)D by just 0.70 nmol/L (from a mean 25(OH)D baseline of 70.3 nmol/L) (29). Finally, using dietary intake of vitamin D to assess status also fails to take into account endogenous production, so exposures are misclassified and probably underestimated.

An EPIC specific limitation of using dietary intakes as a surrogate indicator of vitamin D status is that detailed dietary supplement intake information from the complete cohort is unavailable. More comprehensive data are available from a subset of the cohort (n=36,034) who had additional intake measurements collected via 24-hour recalls. This information highlighted the importance of dietary supplements, particularly in northern European participants (177). For instance, the highest intakes were found amongst participants from Denmark where 51.8% of men and 65.8% of women reported consuming dietary supplements the previous day. Similar high levels were reported amongst participants from Sweden, UK, and Norway. Vitamin D was a frequently used ingredient within supplements

consumed in these countries, partly because cod liver oil intake is often recommended to prevent vitamin D deficiency, as endogenous production is relatively low (177). This higher recorded consumption of dietary supplements for northern European countries is also probably a contributory factor to latitude being such a poor indicator of vitamin D status as was revealed during the derivation of the predictor 25(OH)D scores. For instance, women from Denmark had the highest mean circulating 25(OH)D across all countries.

A more general limitation of all epidemiological dietary analyses, which may have also contributed to the null results observed, is dietary assessment measurement error (random and systematic). This may have caused the dietary intake information collected from participants at baseline to be under- or over-reported leading to imprecise risk estimates and ultimately regression attenuation (178).

In summary, null associations were observed for dietary vitamin D intake and all-cause mortality, cancer mortality and cause-specific mortality. Within EPIC, recorded dietary vitamin D intakes were low, dietary supplement intakes were not recorded, and endogenously produced vitamin D was not considered. Taken together, this suggests that within EPIC, the use of dietary intake as a surrogate measure of vitamin D status is inadequate.

4.2   Derivation and validation of predictor 25-hydroxyvitamin D scores

In this analysis, a sample of participants with serum measurements available from the EPIC study were used to derive and validate circulating 25(OH)D predictor scores. Previously, predictor scores have been successfully derived and applied to assess cancer risk and type 2 diabetes risk in the HPFS (52) and Framingham Offspring (165) cohorts respectively. The EPIC derived models were created by secondary analysing data from previous nested case-control studies (colorectal cancer, prostate cancer, lymphoma and breast cancer). The final predictor scores provided poor estimates of absolute 25(OH)D values, but were reasonably successful at ranking participants into high and low predicted 25(OH)D categories. Consistent results were also yielded when actual and predicted circulating 25(OH)D measurements were used for analyses of colorectal cancer incidence and prostate cancer incidence. Taken together, this suggests that the derivation of predicted 25(OH)D scores may be a practical cost effective approach to be used in chronic disease epidemiological research, especially as the cost of measuring actual circulating 25(OH)D levels may be prohibitively expensive.

4.2.1   Predictor circulating 25-hydroxyvitamin D score derivation

The first stage in creating the EPIC based models was to assess which available variables correlated with circulating 25(OH)D as these would form the basis of the predictor scores. The correlates/determinants included within the predictor scores were: age at blood collection, dietary vitamin D intake, waist circumference, physical activity index, month of blood collection, country of residence, smoking status and intensity, ever use of menopausal hormone therapy, and source study and batch of serum samples. Overall the predictor scores explained 34% and 28% of the variance (R^2) in circulating 25(OH)D amongst men and women respectively, therefore limiting their ability to classify participants' absolute vitamin D status. However, these R^2 values are consistent with previous studies which have investigated correlates of circulating 25(OH)D levels (ranging from 0.21 – 0.42) (50;164;165;179;180).

Generally, the correlates/determinants included in the EPIC derived scores were similarly related to circulating 25(OH)D as per previous studies. Waist circumference (the marker of adiposity used) was inversely correlated with circulating 25(OH)D in both men and women (50). Dietary vitamin D intake was positively correlated with circulating 25(OH)D in both men

and women (49;52;165). The monthly variation in circulating 25(OH)D levels was consistent with previous analyses, with the highest levels found in the late summer/early autumn months and the lowest levels in late winter/early spring months (164;165).

The relationships between mean circulating 25(OH)D and country indicated that latitude is a poor predictor of vitamin D status when making inter-country comparisons across Europe. For women, higher mean serum 25(OH)D levels were found amongst participants from Denmark (latitude ~56°N), with women from Greece (37°N, Athens) and Italy (36°N - 45°N) having among the lowest levels. Men from Italy were found to have the highest mean levels of circulating 25(OH)D, but individuals from the UK (51°N - 52°N) and the Netherlands (52°N) had higher mean 25(OH)D concentrations than participants from Spain (37°N - 43°N). This positive association between latitude and 25(OH)D levels across European countries has been observed in previous analyses. One study of European populations reported that people from Scandinavian countries had the highest mean levels of serum 25(OH)D (181). Similarly, a recent systematic review of elderly participants across Europe calculated a mean increase in serum 25(OH)D of 11.8 nmol/L was associated with a 10 degree increase in latitude of country of residence (47). Potential explanations for this seemingly counter-intuitive relationship between vitamin D status and latitude are that dietary supplement intakes are higher amongst northern European populations (177). Another potential reason may be that northern European populations with lighter skin colour are more inclined to sun expose than southern European populations who have darker skin and less desire to be in the sun for long time periods (47).

Physical activity index, included in the predictor models as a proxy measurement of sun exposure, was positively correlated with circulating 25(OH)D levels, and this was in line with previous research (50;52;164). Age at blood collection was inversely associated with circulating 25(OH)D for women, which is consistent with previous research (3;53;165); but unexpectedly, age was positively correlated with 25(OH)D for men. Current smokers had lower mean circulating 25(OH)D levels than never smokers (49;165). Both age at blood collection and smoking status and intensity were important correlates/determinants of circulating 25(OH)D levels, and as a consequence were included in the predictor scores as covariates; however, both are also prominent risk factors/confounders for chronic disease, so as a consequence, these variables were excluded from the predictor scores when applied to the validation datasets to minimise the risk of statistical over-adjustment.

4.2.2 Predictor circulating 25-hydroxyvitamin D score validations

Correlations between predicted 25-hydroxyvitamin D and actual circulating 25-hydroxyvitamin D

For the EPIC derived predictor scores, relatively low Spearman and Pearson correlation coefficients (ranging from 0.19-0.21) were observed when levels of predicted and actual circulating 25(OH)D were compared in the independent validation datasets. The correlation coefficients from previous predictor scores, when predicted and actual serum levels of 25(OH)D were compared in independent datasets, have ranged from 0.23 – 0.40 in the NHS and HPFS studies, to 0.45 in the WHI, and 0.51 in the Framingham Offspring cohort (50;164;165).

Assessing the validity of predictor scores by calculating the correlation with single measurements of actual circulating 25(OH)D may not be appropriate. Single 25(OH)D measures are influenced by recent sun exposures and behaviours (e.g. beach holiday and sunbathing habits), and are therefore a better indicator of short-term exposures. Intra-class correlation coefficients between circulating 25(OH)D taken 2-3 years apart are around ~0.70 (52;162); however, over longer time periods (5 years plus) the correlations have been shown to decrease ~0.50 (163;164). In chronic disease epidemiology, estimates of long term average exposures - rather than short-term - are required to assess associations. Due to the within-person variation in actual circulating 25(OH)D levels, the use of a single measure as the "gold standard" or "truth" to assess the validity of the predictor scores may not be optimal. The use of such an "alloyed" or imperfect "gold standard" may have meant that the correlation coefficients between predicted and actual circulating 25(OH)D levels in the validation datasets were underestimated by the random within-person error intrinsic to both of these measures (164;182;183).

Actual 25-hydroxyvitaminD measurements by decile of predicted 25-hydroxyvitamin D

An alternative approach to assess the performance of predictor scores is to compare the actual mean circulating 25(OH)D levels by decile of predicted 25(OH)D. While single measures of circulating 25(OH)D are subject to random within-person error, the average of these measurements in each decile should be unbiased and provide population level means (164;178). Within the EPIC validation datasets, mean circulating 25(OH)D levels generally grew with increasing decile of predicted 25(OH)D scores (*P*-trends <0.0001). In men and women, the differences in actual 25(OH)D levels between the extreme deciles were 23.3

nmol/L and 15.2 nmol/L respectively. Within the NHS and HPFS cohorts similar differences (ranging from 21.7-30.7 nmol/L) were reported (52;164).

Cross-classifications of participants by predicted and actual circulating 25-hydroxyvitamin D categories

For both sexes, the ability of the models to classify men and women in the validation datasets into equivalent clinically defined actual and predicted 25(OH)D categories was poor. When three categories of 25(OH)D were used (<50 nmol/L deficient, 50-<75 nmol/L insufficient, and 75+ nmol/L sufficient), the predictor scores cross-classified the vast majority of men in the actual 'deficient' category into the equivalent predictor score derived group (85.4%). However, the predictor score was far less successful at classifying men with higher levels of actual circulating 25(OH)D into corresponding categories, and did not classify any individuals with actual 25(OH)D levels over 75 nmol/L (the optimal level for bone health) into the equivalent predicted category. For women, the predictor score was also poor at cross-classifying into equivalent actual and predicted clinically defined categories. Once more, at actual levels over 75 nmol/L, just 2.4% of women were classified into the corresponding predicted 25(OH)D category. That the predictor scores performed so poorly when estimating absolute vitamin D levels and classifying participants into clinically relevant categories is unsurprising as 66% and 72% of variance in circulating 25(OH)D levels were unexplained for men and women respectively. The WHI derived predictor score also performed poorly when classifying women into clinically defined 25(OH)D categories (50).This indicates that 25(OH)D predictor scores would not have utility in a clinical setting as absolute levels cannot be estimated. Similarly, the predicted scores could not be used in epidemiological studies when clinically defined exposure categories are used. Thus, they are not useful when investigating the elevated risks of all-cause mortality (and other cancer incidence end-points) previously observed at circulating 25(OH)D levels ranging from >93-140 nmol/L when compared against the mid-range reference categories (33-35).

However, within epidemiological studies, rather than absolute values, ranking participants into high and low exposure categories is usually of interest (178). For quintiles, the predictor scores performed better at classifying individuals into equivalent predicted and actual circulating 25(OH)D categories. Across all quintiles, 27% of men and 24% of women were classified into identical predicted and actual 25(OH)D categories. Just under half of men (46.3%) and women (45.8%) were classified into equivalent or parallel quintiles of actual and predicted 25(OH)D. Within the NHS and HPFS derived predictor scores similar proportions of participants were classified into corresponding actual and predicted 25(OH)D quintiles

(164). Also, similarly to these cohorts, the EPIC derived scores classified just 5% of all participants into the extreme opposite quintiles according to actual and predicted 25(OH)D levels (164).

Assessment of colorectal cancer and prostate cancer risk using the predictor circulating 25-hydroxyvitamin D scores

The final validation stage for the EPIC derived predictor scores was to use them to assess risks of colorectal cancer incidence (in the nested case-control and full EPIC datasets) and prostate cancer incidence (in the full EPIC dataset). For both cancers nested case-control studies have been published which measured actual circulating 25(OH)D levels (67;97).

Firstly, the predictor scores were applied to the colorectal cancer nested case-control dataset. Importantly, to ensure independence, the predictor scores for this particular validation stage were derived excluding the actual 25(OH)D samples sourced from the published colorectal cancer analysis (67). In the multivariable model of the earlier published EPIC analysis which measured actual 25(OH)D measurements, a 38% reduced colorectal cancer incidence risk (OR 0.62, 95% CI: 0.47-0.81) was observed when the highest and lowest quintiles were compared (67). When quintiles of predicted 25(OH)D were inputted into the multivariable model, a similar, albeit slightly stronger inverse association was observed (Q5 vs. Q1, OR 0.50, 95% CI: 0.26-0.97; *P*-trend 0.03). This inverse association ceased to be significant when the multivariable model was additionally adjusted for physical activity index, although a significant inverse association remained in the continuous model. Physical activity index was an important correlate/determinant of circulating 25(OH)D included in the predictor scores. Therefore predicted 25(OH)D levels would already have taken into account participant physical activity levels, and including again in the multivariable models when assessing disease risk relationships may be statistical over-adjustment. However, for information purposes, risk estimates for the multivariable models plus physical activity adjustment are presented. Similarly, the multivariable model is also presented minus BMI adjustment. Waist circumference was included as the marker of adiposity in the predictor scores, but due to its close correlation with BMI (r =0.78), additionally adjusting for BMI within the multivariable models when assessing disease risk may also be deemed as statistical over-adjustment. However, it is arguable that biologically these measures are different indicators of adiposity: with waist circumference a better measurement of central obesity which impacts most upon metabolic disorders; while BMI is a more suitable measurement of whole body adiposity. For colorectal cancer incidence, when the multivariable model was not adjusted for BMI, a slightly stronger inverse association was

observed. Colorectal cancer is a suitable end-point to test the sensitivities of these potential over-adjustments as physical activity index and BMI are accepted risk factors for the disease (184-187).

Next, the predictor scores were applied to assess colorectal cancer incidence and prostate cancer incidence in the full EPIC cohort. For colorectal cancer incidence, in the sexes combined multivariable model, once more a significant inverse association was observed (Q5 vs. Q1, HR 0.75, 95% CI: 0.62-0.91; *P*-trend 0.009). For prostate cancer incidence, a previous EPIC nested case-control study reported a non-significant positive association in the multivariable model (Q5 vs. Q1, OR 1.28, 95% CI: 0.88-1.88; *P*-trend 0.19) (97). When the men's predictor 25(OH)D score was applied to the full EPIC cohort, a non-significant positive association was also observed (Q5 vs. Q1, HR 1.08, 95% CI: 0.90-1.39; *P*-trend 0.70).

Overall, encouragingly consistent colorectal cancer and prostate cancer associations were observed between models which used actual and predicted 25(OH)D measurements. Likewise in the NHS and HPFS cohorts, similar results were yielded between analyses using predicted and actual circulating 25(OH)D measurements for endpoints, including colorectal cancer incidence and mortality, pancreatic cancer incidence, prostate cancer incidence, and hypertension (164). For instance, in a joint NHS and HPFS nested case-control analysis for colorectal cancer incidence, risk estimates of 0.82 and 0.78 were yielded per increment of 25 nmol/L for models using actual and predicted 25(OH)D respectively (52;130;164). This evidence from EPIC and the U.S. cohorts suggests that predicted 25(OH)D may be an acceptable surrogate indicator of actual 25(OH)D levels for epidemiological studies.

Validation of the Health Professionals Follow-Up Study predicted 25-hydroxyvitamin D score

The adaptation and application of the HPFS derived 25(OH)D predictor score (52) within men in EPIC was unsuccessful. In the validation dataset, the correlation coefficients between actual and HPFS predicted 25(OH)D score was just 0.05. Importantly, where the EPIC derived scores performed well (actual mean per decile of predicted 25(OH)D and the duplication of the previous inverse colorectal cancer incidence association), the HPFS predictor scores performed poorly. For instance, when the adapted HPFS score was applied to the men's EPIC dataset, a non-significant positive association was observed when the highest and lowest quintiles of predicted 25(OH)D were compared (HR 1.23, 95% CI: 0.76-

1.99). Bertrand *et al.,* (164) had proposed that 25(OH)D prediction scores may be used across different cohorts as long as participants within each population were relatively homogenous for characteristics, demographics and residential latitude locations. HPFS participants were not located at latitudes above 47°N which meant that EPIC participants in northern France, Germany, the Netherlands, UK, and Denmark had to be excluded when the predictor score was adapted. Men within the HPFS are, due to all working within the same sector, probably similarly educated and from the same or similar social class. In contrast, the EPIC cohort is heterogeneous with regards to occupations and social class (as assessed by the proxy measurement of highest educational level attained). Finally, inter-cohort differences in exposure measurement instruments, and the associated systematic and random measurement error, may have also contributed to the failed adaption of the HPFS predictor score within EPIC. However, although the adaptation and application of the HPFS derived 25(OH)D score within EPIC was unsuccessful in this instance, predictor scores may be transferable where inter-cohort differences are minimal.

4.2.3   Strengths and limitations of using predictor circulating 25-hydroxyvitamin D scores

Strengths of using predictor circulating 25-hydroxyvitamin D scores

Practical and cost effective analytical approach

Within epidemiological studies, 25(OH)D prediction scores are a practical and cost effective alternative to investigate vitamin D-chronic disease associations. Actual 25(OH)D measurements are required from a sample of a cohort and then, providing the information on correlates/determinants is available, can be used to derive predictor scores. While using actual circulating 25(OH)D measurements to assess disease risk is viewed as the "gold standard" approach, the high laboratory costs often mean that this is prohibitively expensive. Instead the larger cohort studies have generally used the nested case-control study design to investigate vitamin D-chronic diseases associations. Although costs are minimised by analyses including a subset of cohort participants, the nested case-control datasets created are specific for a disease or disease subtype; as the control participants, usually selected using incidence density sampling, are uniquely matched to cases by follow-up time, time of year of blood collection, age, and other criteria. This means that the relationships between vitamin D and other disease end-points cannot be carried out without creating a new disease-specific nested case-control dataset. The advantage of using predicted 25(OH)D

scores is that the disease risk analyses can include all cohort participants (minus those who were used in the derivation of the predictor score) and multiple end-points can be studied. A further advantage is that the scores can be re-derived to provide updated predicted 25(OH)D estimates as additional exposure information is collected participants during the follow-up period.

Limitations of using predictor circulating 25-hydroxyvitamin D scores

Large amount of variability in circulating 25-hydroxyvitamin D unexplained

For both men and women's 25(OH)D predictor scores 66% and 72% of variability in circulating 25(OH)D was unexplained respectively. This was probably due to measurement error in the correlates/determinants of 25(OH)D included in the predictor scores, plus known determinants of vitamin D status - which were unmeasured within EPIC – excluded from the predictor scores. For instance, vitamin D supplement intakes were unknown, and the European inter-country positive association between latitude and vitamin D status suggests that this is an important correlate/determinant of 25(OH)D levels. Additionally, information on race/ethnicity of EPIC participants, an important determinant of vitamin D status (9), has not been collected. Although the assumption is that the vast majority of participants are Caucasian, accurately capturing this information may have explained more variance in the predictor model. The inclusion of genetic information within the predictor scores may also have explained more variance as results from twin studies are suggestive of a significant hereditable component in circulating 25(OH)D levels (188). In particular, certain SNPs related to vitamin D binding protein (VDP) gene, such as rs4588 and rs7041 have been consistently associated with circulating 25(OH)D levels (180;189).

Within the predictor scores, physical activity index was used as a proxy measurement for actual UV exposure. The physical activity index variable measured walking, cycling, gardening, sports, DIY, stair climbing, and occupational activity. Many of these activities may be performed indoors, so physical activity in the EPIC models is probably an inadequate proxy variable of sun related behaviour. Also, the amount of vitamin D endogenously produced while outside is dependent on many factors, including time spent outside, clothing worn, sunscreen use, and time of day and season of exposure. All of this important information has not been captured within the physical activity index variable and consequently the predictor models.

Different laboratories and 25-hydroxyvitamin D assays used

The actual circulating 25(OH)D samples used to derive and validate the predictor scores were sourced from four different nested case-control studies (colorectal cancer, prostate cancer, lymphoma, or breast cancer). This meant that 25(OH)D levels were assessed at different times and in four different laboratories. Furthermore, across the four studies, two different assays were used. This may have introduced extraneous inter-study laboratory and batch variation into the circulating 25(OH)D measurements. In an attempt to partially adjust for this methodological limitation, a variable for which study and batch the circulating 25(OH)D data were sourced from was additionally included in the predictor scores. Inclusion of this source study-batch variable explained 7% of the circulating 25(OH)D variance in men and women. Lips et al., (181) recommend that the optimal method of control would be to cross-calibrate samples by analysing a selection of the same serum data at different laboratories. This was not possible in this EPIC study as the serum samples had been measured prior to this current analysis.

Confounding and over-adjustment by the correlates/determinants of 25-hydroxyvitamin D

An objection raised to the use of vitamin D predictor scores is that of confounding and/or statistical over-adjustment, as some of the correlates/determinants of 25(OH)D are also important chronic disease risk factors (e.g. age, smoking status, waist circumference and physical activity) (50;164). This limitation is also applicable to actual circulating 25(OH)D measurements which intrinsically incorporate these factors. For instance, within EPIC, higher levels of actual circulating 25(OH)D were correlated with being physically 'active', not smoking and being slimmer. An advantage of predictor 25(OH)D scores is that sensitivity analyses can be performed to assess the influence of possible statistical over-adjustment. During the current study's validation stage sensitivity analyses, no evidence of statistical over-adjustment was revealed; although this does not mean that this bias can be entirely discounted. For prudence, to evaluate the impact of possible statistical over-adjustment, multivariable risk estimates minus BMI adjustment, and with additional adjustment for physical activity index are presented.

Lack of generalizability to other populations

As the analysis of the HPFS predictor score within EPIC demonstrated, 25(OH)D predictor scores created within a particular population may only be valid for use within the same population. However, this lack of external generalizability should not affect the internal validity of the predictors score creation, validations and application to assess disease risk.

Small amount of circulating 25-hydroxyvitamin D samples available for predictor score development and validation

EPIC is a large and heterogeneous cohort including 521,448 participants from 23 centres in 10 European countries. This current study used serum samples from just 6,118 participants (1.2%) to derive and validate the 25(OH)D predictor scores. Additional 25(OH)D samples may better capture the heterogeneity of EPIC participants and ultimately improve the performance of the predictor scores. As additional samples do become available from future nested case-control analyses, the predictor scores could be re-derived and validated once more.

4.2.4 Conclusion

For the first time in a European population predicted circulating 25(OH)D scores were derived and validated. Consistent with previous studies in the U.S. the predictor scores provided a poor indication of absolute vitamin D status and would therefore not be useful in a clinical setting or when clinically relevant categories are used in observational research. However, for epidemiological research, where ranking of participants is often sufficient, the predictor scores may have utility. Encouragingly, the predictor scores were able to replicate results from previous colorectal cancer incidence and prostate cancer incidence nested case-control studies which used actual circulating 25(OH)D measurements. This suggests that predicted 25(OH)D scores may be a practical cost-effective alternative to measuring actual vitamin D levels. However, when using predictor scores, sensitivity analyses should always be undertaken to ensure that bias is not introduced due to confounding and statistical over-adjustment by the correlates/determinants of 25(OH)D also being confounders in the disease risk analyses.

4.3 Predicted 25-hydroxyvitamin D levels and cancer incidence and mortality, all-cause mortality and cause-specific mortality

In this prospective analysis within the EPIC cohort, the validated predicted 25(OH)D scores were applied to assess the relationships with cancer incidence and mortality, circulatory disease mortality, respiratory disease mortality, digestive disease mortality, and all-cause mortality. This was the first time that predictor 25(OH)D scores have been used in European populations to assess disease incidence and mortality risks. As was revealed during the predictor scores validation stage, an inverse association was observed for colorectal cancer incidence, a result consistent with the majority of previous studies. Additionally, inverse predicted 25(OH)D score associations were observed for: overall cancer incidence and mortality; lung cancer incidence and mortality; kidney cancer incidence; stomach and oesophageal cancer incidence; pancreatic cancer incidence and mortality; thyroid cancer incidence; prostate cancer mortality; all-cause mortality; circulatory disease mortality; respiratory disease mortality; and digestive disease mortality. Many of these end-points have not previously been associated with vitamin D. The observed results were generally consistent across strata of other chronic disease risk factors and stable when cases/deaths recorded within the first 5 years of follow-up were excluded. However, due to the methodological limitations of 25(OH)D predictor scores, and observational epidemiology in general, it is important to acknowledge that alternative explanations may explain some or all of these observed relationships.

4.3.1 Cancer incidence and mortality

Overall cancer

To date, only a handful of prospective studies have investigated the vitamin D-cancer incidence relationship. The inverse relationship observed in this current EPIC analysis is consistent with a HPFS analysis which also used predicted 25(OH)D levels (52). Of the studies which measured actual circulating 25(OH)D levels, a recent German cohort analysis observed a reduced cancer incidence risk amongst men in the highest exposure group of actual circulating 25(OH)D when compared with the lowest group; although no association was observed for women (57). Similarly, a small Swedish cohort also observed increased

cancer incidence risk at lower levels; however, this analysis also revealed an elevated risk amongst individuals with actual 25(OH)D levels >98 nmol/L when compared against the mid-range reference category (vs. 46-93 nmol/L, RR 1.68, 95% CI: 1.06-2.65) (35). Due to the predictor scores estimating absolute 25(OH)D levels so poorly (i.e. being unable to adequately classify individuals with actual circulating 25(OH)D levels >75 nmol/L into the equivalent predicted 25(OH)D category), an investigation of the elevated cancer incidence risk observed amongst individuals with higher 25(OH)D levels was not possible.

When the predicted 25(OH)D levels-overall cancer incidence association was analysed by strata of lifestyle, demographic, anthropometric, and dietary variables, non-significant interactions were observed for all of the various factors considered. However, no association was observed amongst never smokers, which suggests that residual confounding may have influenced the vitamin D-overall cancer incidence relationship. To further investigate this possible bias, cancer incidence and mortality cases were split into smoking and non-smoking related cancer cases. For smoking related cancer incidence, a stronger inverse association was observed than for overall cancer incidence (a 35% lower risk per 25 nmol/L increment in predicted 25(OH)D vs. 13% lower risk for overall cancer incidence). Importantly, significant inverse associations for smoking related cancer incidence were observed across all strata of smoking status, including for never smokers. Despite this, residual confounding by smoking habits cannot be ruled out as an explanation for the inverse cancer incidence and mortality associations observed as tobacco consumption is a major cause of cancer.

For cancer mortality, a stronger inverse relationship was observed than for overall cancer incidence (35% lower risk per 25 nmol/L increment in predicted 25(OH)D vs. 13% lower risk). In the HPFS study, Giovannucci *et al.,* (52) reported similar cancer incidence and mortality associations to those observed within EPIC. However, mixed results have been reported in the previous studies which measured actual circulating 25(OH)D levels to assess cancer mortality risk. Three previous studies have reported elevated risks in the highest exposure group when compared against the mid-range or lowest group (35;36;54). Freedman *et al.,* (36) in an NHANES III analysis reported an increased cancer mortality risk (1.9-fold) amongst men with actual 25(OH)D levels >100 nmol/L when compared against the <37.5 nmol/L exposure group. While, the aforementioned Swedish cohort, observed a 2.5-fold increased cancer mortality risk (similarly to what they observed for cancer incidence) amongst the >93 nmol/L actual 25(OH)D group compared to the mid-range group (<39 nmol/L) (35). From the remainder of previous studies, some reported inverse associations (59-61), but most reported non-significant relationships (62-65;156). Overall, whether vitamin

D has a beneficial biological role for overall cancer incidence and mortality remains unclear. Heterogeneous results have been observed, with unexpected positive associations observed amongst individuals with actual circulating 25(OH)D levels >93/100 nmol/L. The lack of consistency between individual studies of overall cancer incidence and mortality may be a consequence of the considerable aetiological heterogeneity of individual cancers (and indeed within subtypes of the same cancers).

Giovannucci *et al.,* (52), based upon ecological data, hypothesised that any biological role for vitamin D on cancer would be stronger for tumours sited in digestive tract. Results from the HPFS study were supportive of this with a 48% and 55% lower digestive tract cancer incidence and mortality association observed respectively per 25 nmol/L increment of predicted 25(OH)D level (52). Similarly to this result, within EPIC, stronger associations were observed for digestive system cancer incidence than for non-digestive system cancer incidence, although significant risk estimates were observed for the latter. The characteristics of digestive system cancers that may make them more responsive to the biological effects of vitamin D are currently unknown.

Colorectal cancer

As was revealed during the predictor score validation stage, an inverse association was observed for colorectal cancer incidence. This result is consistent with the vast majority of prospective research (67;70;71;77). Of all the cancers, only for colorectal cancer does an anti-carcinogenic role for vitamin D seem persuasive. However, as these results are all observational, intervention studies are required to determine if this relationship is causal. Previous research investigating the relationships between circulating 25(OH)D and colorectal cancer mortality have only occasionally been conducted. In two of these studies (one from EPIC), which included individuals with colorectal cancer only, higher pre-diagnosis levels of actual circulating 25(OH)D were associated with lower risks of colorectal cancer deaths (83;85). Within this current EPIC analysis, which included all participants (rather than solely individuals with colorectal cancer), an inverse relationship was observed, although this was non-significant. Whether actual or predicted pre-diagnosis estimates of circulating 25(OH)D are the optimal measure for assessing mortality is uncertain. This is because behaviours may change after a cancer diagnosis, and such modifications would impact upon the correlates/determinants of vitamin D and ultimately predicted/actual 25(OH)D levels. Within EPIC, post-diagnosis lifestyle, anthropometric and dietary information has not been collected from participants. If in the future this information becomes available, the predictor

scores could be re-derived to estimate post-diagnosis 25(OH)D levels and use this to re-assess mortality risks.

Lung cancer

The inverse association observed for lung cancer incidence was stronger and only significant among women. Previous prospective analyses of actual or predicted 25(OH)D have reported non-significant associations (52;126;127). The observed inverse association should, however, be interpreted with caution due to the sensitivity of adjusting for BMI in the multivariable models. When BMI was removed from the model, the previously inverse association became null (HR per 25 nmol/L increment in predicted 25(OH)D, 1.01, 95% CI: 0.81-1.26). For predicted 25(OH)D levels and lung cancer mortality, a significant inverse association was observed in the sexes combined categorical model. A previous NHANES III analysis observed no association for actual 25(OH)D and lung cancer mortality (128). Similarly, to the lung cancer incidence results, this inverse association with lung cancer mortality was no longer present when the multivariable model was not adjusted for BMI. Overall, more large studies investigating the vitamin D-lung cancer relationship are warranted. In particular, studies which are suitably powered to investigate the association by smoking status are required.

Kidney cancer

This was the largest prospective study to date (in terms of cases) to investigate the vitamin D-kidney cancer incidence relationship. The significant inverse association observed was contrary to the null results observed in the VDPP nested case-control analysis which measured actual circulating 25(OH)D levels (129). However, the result was consistent with a recent HPFS/NHS analysis, in which a 25 nmol/L increment in predicted 25(OH)D was associated with a 44% lower kidney cancer risk (versus a 53% lower risk within EPIC) (130). In the HPFS/NHS analysis similar strength inverse associations were observed amongst men and women. Within EPIC, the inverse association was stronger for men; however, although non-significant, an inverse association was also observed for women. The inconsistent results between the VDPP analysis which measured actual levels of 25(OH)D, and the NHS/HPFS and EPIC analyses which used predicted 25(OH)D levels, could be because single actual 25(OH)D measurements may not reflect long-term exposure due to the within-person variation in vitamin D levels. However, whether predicted 25(OH)D estimates provide a more stable long-term indicator of vitamin D status than actual measurements in EPIC is unknown.

Stomach and oesophageal cancers

The 40% lower risk observed per 25 nmol/L increment in predicted 25(OH)D should be interpreted cautiously as this association weakened and became non-significant when the multivariable model was not adjusted for BMI. This suggests that the association may be an artefact of confounding and/or statistical over-adjustment. Prospective studies investigating the vitamin D-stomach and oesophageal cancer relationship are rare. In the HPFS study, a significant inverse associations was observed for predicted 25(OH)D with oesophageal cancer, with a non-significant relationship observed for stomach cancer (52). In a VDPP nested case-control study, null results were reported for oesophageal and gastric cancer when analysed together (131).

Bladder cancer

The null result observed for bladder cancer incidence is consistent with all but one previous prospective study (52;134). Only in the ATBC study was an inverse association observed (133); that participants within this analysis were all men who smoked could have been a factor influencing this result. However as only four prospective studies have been carried out to date, more data for the vitamin D-bladder cancer relationship are required.

Pancreatic cancer

The inverse association observed for predicted 25(OH)D levels with pancreatic cancer incidence is similar to a recent joint HPFS, NHS, PHS, WHI, and WHS nested case-control analysis which measured actual circulating 25(OH)D levels (125). However, this is contrary to the VDPP nested case-control analysis which included data from eight prospective cohorts and reported a twofold elevated pancreatic cancer risk at actual 25(OH)D levels over ≥100 nmol/L when compared against the mid-range category (50-<75 nmol/L) (40). Unfortunately, investigating whether individuals with 25(OH)D levels ≥100 nmol/L was not possible in this current EPIC analysis as the predicted scores were unable to provide adequate estimates of absolute vitamin D levels. For pancreatic cancer mortality, unsurprisingly due to the poor prognosis of patients with the disease, an inverse predicted 25(OH)D association of similar strength was also observed. In general, more research is required to understand role of vitamin D on pancreatic carcinogenesis. In particular, other prospective studies which measure actual 25(OH)D levels are required to further investigate the possible elevated risk at high concentrations.

Liver cancer

To date, no previous prospective studies investigating the vitamin D-liver cancer incidence relationship have been carried out. Within this current analysis, a non-significant inverse association was observed in the multivariable model. However, when the multivariable model was not adjusted for BMI, the inverse association strengthened and became significant. This result is indicative of a possible beneficial role for vitamin D on liver cancer incidence, even though the strength, and whether this relationship was significant, was sensitive to BMI adjustment. Either way, the result highlights a possible novel vitamin D-cancer relationship which warrants further investigation; specifically, with prospective studies which measure actual circulating 25(OH)D levels.

Non-Hodgkin's lymphoma

No association was observed for predicted 25(OH)D and overall NHL incidence. This null result for overall NHL risk is consistent with all previous published analyses, including an EPIC nested case-control study (52;135;136). However, NHL is a heterogeneous disease with multiple subtypes which may have varying aetiologies (190). Within the aforementioned EPIC analysis, a positive B-NHL association was observed amongst those individuals with actual 25(OH)D levels over ≥75 nmol (vs. 50-<75 nmol/L, OR 1.36, 95% CI: 1.00-1.83; *P*-trend 0.007) (136). As the predicted scores were unable to classify individuals with 25(OH)D levels ≥75 nmol, this association could not be investigated in the current analysis.

Brain cancer

A suggestive inverse association was observed for predicted 25(OH)D and brain cancer incidence, with a near significance inverse trend (*P*-trend 0.05) observed for women. The only previous prospective study reported a non-significant positive association using a predicted 25(OH)D score in the HPFS study (52). For brain cancer mortality, no association was observed.

Skin cancer

The positive association observed between predicted 25(OH)D and skin cancer incidence was consistent with a HPFS analysis which also used predicted 25(OH)D levels (52). The vitamin D-skin cancer relationship is likely to be confounded by sun exposure habits.

Detailed information on sun exposure behaviours was not available within EPIC. As sun exposure habits/behaviours would likely be a positive confounder, the observed positive skin cancer association would be expected to disappear or attenuate after adjustment for these factors. However, that a positive association was observed suggests that the predicted 25(OH)D scores may be an acceptable surrogate measure of vitamin D status.

Thyroid cancer

No previous studies have investigated the relationship between vitamin D and thyroid cancer. Therefore the 50% lower thyroid cancer incidence risk (95% CI: 0.50-0.97) observed per 25 nmol/L increment in predicted 25(OH)D levels is a novel finding. However, this association in the sexes combined continuous model should be interpreted cautiously. Firstly, this analysis included just 325 thyroid cancer cases. Secondly, in sensitivity analyses, the association attenuated and became non-significant when the multivariable model was not adjusted for BMI; suggesting that the association may be an artefact of confounding and/or statistical over-adjustment. Despite this, the results highlight a possible role for vitamin D on thyroid carcinogenesis. In this scenario, the predictor scores have acted as a hypothesis generator and flagged up a possible exposure-endpoint relationship where future resources may be focussed; specifically, nested case-control studies which measure actual circulating 25(OH)D levels.

Prostate cancer

As was revealed during the predictor scores validation stage, a non-significant positive association was observed for prostate cancer incidence. The vast majority of prostate cancer nested case-control studies which measured circulating 25(OH)D have yielded null or non-significant positive associations (86-95), although three recent Nordic nested case-control studies have observed significant positive associations (37-39). For prostate cancer mortality, the observed inverse association is consistent with a recent HPFS nested case-control analysis which reported a 57% reduced risk (95% CI: 0.24-0.76) when the highest quartile of pre-diagnosis circulating 25(OH)D was compared with the lowest quartile (102). Within EPIC, a 50% reduced risk was observed for the equivalent comparison of pre-diagnosis predicted 25(OH)D quintiles. However, this inverse association should be interpreted cautiously. First, only 228 prostate cancer deaths were included within the analysis, meaning that this association could be due to chance. Second, residual confounding from unmeasured factors may have impacted upon this association. Specifically, failure to control for screening from prostate-specific antigen (PSA) testing may have biased the results; however within the HPFS analysis, virtually unchanged associations

were observed even after adjustment for PSA screening occurrence and frequency (102). Overall, larger pooled studies with an increased number of recorded deaths, plus information on PSA screening and treatments, are required to confirm the potential beneficial role of vitamin D on prostate cancer mortality.

Breast cancer

Similarly to the recently published EPIC nested case-control analysis using actual circulating measurements (117), a null result was also observed for the predicted 25(OH)D-breast cancer incidence relationship. A similar null result was also observed for breast cancer mortality. All but two (110;111) previous prospective nested case-control studies which measured actual circulating 25(OH)D levels have similarly reported no association for breast cancer incidence (110;112-117). Although a meta-analysis of these studies (excluding the EPIC nested case-control study which was the largest to date) reported an inverse association which was of borderline significance (highest vs. lowest circulating 25(OH)D quintiles pooled OR 0.87, 95% CI: 0.77-0.99) (118). However, when updating the meta-analysis calculation, the inclusion of the EPIC nested case-control study risk estimate (highest vs. lowest 25(OH)D quintiles pooled meta-analysis OR 0.92, 95% CI: 0.80-1.06) or the risk estimate using the predicted 25(OH)D score (highest vs. lowest 25(OH)D quintiles pooled meta-analysis OR 0.92, 95% CI: 0.82-1.02) resulted in this association weakening and losing significance. Overall, the largely null prospective evidence is contrary to case-control evidence which have usually reported inverse associations, most likely due to reverse causality bias. The National Cancer Institute Cohort Consortium of more than 40 studies, is currently undertaking a pooled analysis of the actual circulating 25(OH)D-breast cancer relationship. This analysis will have sufficient power to analyse the associations by breast cancer subtype (ER- and ER+) and by menopausal status, which may provide further insights into the relationship.

Ovarian and endometrial cancers

The null result observed between predicted 25(OH)D levels and incidence of ovarian cancer is consistent with previous prospective analyses (143;144). For endometrial cancer, no association was observed in the multivariable models (HR per 25 nmol/L increment in predicted 25(OH)D, 0.92, 95% CI: 0.63-1.35). However, in sensitivity analysis, when the multivariable model was not adjusted for BMI, a strong significant inverse association was observed (64% lower risk per 25 nmol/L increment in predicted 25(OH)D). BMI is an established risk factor for endometrial cancer with a recent meta-analysis of 24 studies reporting a 1.6-fold higher risk per 5 kg/m^2 increment of BMI (191). This sensitivity of the

predicted 25(OH)D-endometrial cancer relationship to BMI adjustment suggests that the association may be an artefact of confounding and/or statistical over-adjustment. This association should therefore be interpreted cautiously.

4.3.2   All-cause and cause-specific mortality

All-cause mortality

The inverse all-cause mortality association observed for predicted 25(OH)D is consistent with most previous research where similar linear associations have been observed (59;61;63;64;147;149-151). Within this current analysis, a significant linear inverse association was observed for men and women, for all countries, and across strata of all lifestyle, demographic, anthropometric, and dietary variables considered. The observational evidence indicating a beneficial role for vitamin D against deaths from all-causes has been supported by a meta-analysis of 18 smaller trials which reported a significant RR of 0.93 (95% CI: 0.87-0.99) for participants who received vitamin D supplements (153).

Despite this seemingly convincing evidence for a beneficial effect of vitamin D against all-cause mortality, three recent prospective observational studies have reported J- or U-shaped 25(OH)D-all-cause deaths relationships, with increased risks at higher levels (33-35). Within these analyses, elevated all-cause mortality risks were observed for those individuals with actual 25(OH)D levels ranging from >93-140 nmol/L, when compared against mid-range reference groups (33-35). Whether these associations are true or an artefact of other unmeasured criteria remains unclear. Due to the predicted scores yielding poor absolute 25(OH)D estimates, this analysis was unable to investigate whether higher 25(OH)D levels were correlated with higher mortality risks. Larger observational studies which measure actual circulating 25(OH)D are required to investigate any possible adverse effects of higher levels of vitamin D.

Circulatory disease mortality

This was the largest study to date (n=3,641 deaths) to investigate the vitamin D-circulatory disease mortality relationship. Previous prospective studies, which measured actual circulating 25(OH)D levels, have been smaller, with only two having more than 1,000 recorded circulatory disease deaths. Within these studies linear inverse associations have

usually been observed, with a recent meta-analysis of 18 prospective studies reporting a 21% reduced vascular mortality risk when the highest and lowest quartiles of actual circulating 25(OH)D were compared (61). The inverse association observed in the current EPIC study was present for men and women, across all countries and strata of all lifestyle, demographic, anthropometric, and dietary variables considered. The inverse relationship was also similar for ischaemic heart disease and cerebrovascular disease deaths. The totality of the observational evidence is indicative of possible role for vitamin D in reducing circulatory disease deaths. However, these observed inverse correlations require verification in RCTs.

Respiratory disease and digestive disease mortality

Remarkably strong inverse associations were observed for respiratory disease and digestive disease mortality, especially in the continuous models (88% and 89% lower risks observed respectively per 25 nmol/L increment in predicted 25(OH)D). Although for respiratory disease, this association weakened but remained strong and significant when the multivariable model was not adjusted for BMI (62% lower risk per 25 nmol/L increment in predicted 25(OH)D). The inverse respiratory disease mortality relationship is consistent with the three previous prospective studies that have investigated the association using actual circulating 25(OH)D measurements (59;61;62). Similarly, for digestive disease mortality, the inverse association observed was consistent with the one previous prospective analysis, where a 72% reduced risk was observed when the highest and lowest quartiles of actual 25(OH)D were compared (62). Plausible biological mechanisms for vitamin D on digestive and respiratory diseases include: acting as an anti-inflammatory and anti-fibrosis agent; plus elevated antiviral response (158;192).

Despite this biological plausibility, other explanations for the inverse associations observed should be considered. For instance, reverse causality may explain part of the inverse associations observed. The results may be biased by pre-clinical disease at baseline influencing the correlates/determinants of 25(OH)D included within the predictor scores. The result of which may be an artificially elevated disease risk observed amongst individuals with lower predicted 25(OH)D levels. However, in sensitivity analyses, similar strength associations were observed when individuals whose follow-up time was less than 5 years were excluded. While when the analyses were limited to individuals who self-reported being in 'excellent' or 'good' health at baseline, similar inverse associations were observed. Thus, although part of the inverse associations may be caused by reverse causality, the sensitivity analyses conducted were unsupportive of this possibility. A further explanation for the

unrealistically strong inverse associations observed maybe that the predictor scores provide an indicator of 'good general health' rather than estimating vitamin D levels. This is because having higher predicted 25(OH)D levels was associated with having a lower BMI, drinking less alcohol, and being younger and physically active. However, this supposition is also relevant for actual 25(OH)D measurements which also intrinsically incorporate these factors.

Overall, few previous studies have investigated the vitamin D-respiratory disease and digestive disease mortality associations. The results from this current analysis are indicative of strong biological roles for vitamin D on respiratory disease and digestive disease mortality. Further large studies with sufficient cases and longer follow-up times are required to validate and scrutinise these relationships.

4.3.3   Strengths and limitations

Strengths

Large prospective cohort with low losses to follow-up

This was the largest study to date to derive and apply predicted 25(OH)D scores to investigate the vitamin D-cancer incidence and mortality, all-cause mortality and cause-specific mortality relationships. The large number of participants and recorded endpoints (cases/deaths) meant that these associations could be investigated thoroughly. Participants were predominantly sourced from the general population, and comprehensive cohort follow-up by each of the participating study centres ensured low losses to follow-up (<2%) (193).

Variability and detail of exposure information

The current analysis contained participants from eight European countries, within which wide exposure ranges for dietary and lifestyle variables were recorded. Participants were resident in northern to southern European countries at differing latitudes, and with heterogeneous dietary and lifestyle habits. At baseline extensive dietary, lifestyle, demographic, health, and anthropometric information was collected from all participants (168). For this analysis it meant that thorough examination of all possible factors which may confound predicted 25(OH)D levels and the multiple end-points considered could be investigated.

Reverse causality

Participants experiencing pre-clinical disease symptoms at baseline may have had as a consequence a lower waist circumference/BMI, lower dietary intakes and physical activity levels. As a result, such individuals would have lower predicted 25(OH)D levels, thus artificially increasing the chronic disease incidence/mortality risks amongst these participants. However, participants who self-reported previous ill-health at baseline (i.e. cancer, heart disease, stroke, and diabetes) were excluded from the mortality analyses. Further, similar all-cause and cause-specific mortality associations were observed when the analyses were limited to those individuals who self-reported being in 'excellent' or 'good' health at baseline. Finally, all of the observed associations remained stable after cases/deaths recorded within the first five years of follow-up were excluded. Taken together, although reverse causality cannot be ruled out the sensitivity analyses revealed no evidence that the observed associations were caused by this.

Residual confounding

Although detailed information on possible confounding factors was collected from all participants at baseline, as with all epidemiological research, the observed relationships could be biased by residual confounding from unmeasured/poorly measured factors, or the inability of the statistical models to capture complex interrelationships between exposure variables. For instance, information on cancer screening was unavailable, which may have meant that the observed inverse cancer mortality associations are merely an artefact of individuals with higher predicted 25(OH)D levels undergoing screening earlier, and as a consequence early stage treatable tumours being identified.

Predicted vitamin D scores may also be especially sensitive to residual confounding and/or statistical over-adjustment as the correlates/determinants of 25(OH)D used at the derivation stage may also be confounders when assessing the risks of chronic disease incidence and mortality. Within this analysis, separate risk estimates were presented for the multivariable model minus BMI adjustment, plus physical activity index adjustment. For the results where large disparity between models with or without these adjustments was identified (lung cancer incidence and mortality; stomach and oesophageal cancer incidence; liver cancer incidence;

thyroid cancer incidence; endometrial cancer incidence; and respiratory disease mortality), the observed associations should be interpreted with caution due to this instability.

Different exposure assessment methods used

The multi-centre design of EPIC meant that different exposure assessment methods were used. For instance, diet was measured face-to-face in Greece using a quantitative dietary assessment questionnaire with 254 food items (168). In contrast, in Norway diet, was measured by a self-administered semi-quantitative food frequency questionnaire containing 88 food items (168). Similarly, there were differences between EPIC centres in the measurement of anthropometric measurements. In the majority of centres a similar protocol was used as anthropometric measurements were measured at baseline. However, in the France and Oxford (UK) centres these measurements were self-reported (168). Although substantial efforts have been undertaken to standardise exposures across countries/centres (170;194), the possibility exists that extraneous measurement error may have been introduced by the variable methods used.

End-point information collected differently across countries/centres

Cancer registries are used to identify cases during follow-up in Denmark, Italy, the Netherlands, Norway, Spain, Sweden and the UK (168). In the remaining countries (France, Germany and Greece) a combination of methods are used to identify cases, including: health insurance records, cancer and pathology registries and through active follow-up of participants and their next-of-kin (168). These differences in recording cancer cases may have meant that diagnoses may be underreported in some countries which may have influenced the observed results.

Multiple end-points and chance associations

Within this study 36 incidence and mortality end-points were considered (not including subgroup analyses across strata of other dietary, lifestyle, demographic and anthropometric factors). Due to these multiple analyses several false-positive or chance associations would be expected to be observed (*P*-value of <0.05).

Generalizabilty to non-Caucasian populations

Although this information was not formally collected, EPIC is thought to consist of Caucasian participants only. This means that the observed associations may not be applicable to other

racial/ethnic groups. However, this also means that the study results are unlikely to be confounded by ethnicity/race, which is an important determinant of vitamin D status.

Post-diagnosis exposures unknown

For the cause-specific mortality analyses, the derived scores were based on pre-diagnosis information, meaning that any post-diagnosis changes in dietary and lifestyle habits were unknown. Within the predictor scores, pre-diagnosis physical activity index (included in the predictor models as a proxy measurement of sun exposure) was an important correlate/determinant of circulating 25(OH)D levels. Evidence from the UK and U.S. has shown that physical activity levels reduce in cancer patients post-diagnosis (195;196). Such a reduction would be expected to lower predicted 25(OH)D levels, although whether such changes occurred within EPIC is unknown.

Information on post-diagnosis therapeutic treatments is also unknown in EPIC. The possibility exists that individuals with higher predicted 25(OH)D levels may have been more likely to undergo effective treatment earlier, and as a consequence had better disease prognosis and a lower mortality risk. Within EPIC, higher predicted 25(OH)D levels were correlated with younger age and higher educational level attained (in women). For cancer, evidence has shown that the likelihood of individuals to accept screening and chemotherapy treatment is higher amongst those with higher incomes, educations and social class (197;198). Furthermore, older age has previously been associated as a risk factor for a poorer standard of cancer treatment (197).

Across the multiple European countries included within EPIC, clinical practices and treatments may have differed which may have introduced bias into the observed results. In an attempt, to partially adjust for this inter-country heterogeneity, all models were stratified by study centre. Similarly, during the 12 years of follow-up time treatments may also have changed within countries; although the mortality associations observed were generally stable throughout the follow-up period.

Subtypes of diseases not considered

The endpoints considered were for overall incidence/mortality for a certain type of cancer/chronic disease. This is likely to be an oversimplification, as increasing aetiological heterogeneity has been identified between subtypes of the same cancer. For instance,

previous EPIC analyses have often reported heterogeneous associations with various risk factors by colorectal cancer sub-site. For example, reduced risk of proximal colon cancer has been observed amongst physically active participants when compared against inactive participants; but no association was observed for distal colon and rectal cancers (199). Similarly, an inverse dietary fibre association was observed for proximal colon and rectal cancers, but not distal colon cancers (200). For breast cancer, a recent pooled analysis which included just under 1 million women reported inverse fruit and vegetable associations for ER-, but not ER+ tumours (201). The analyses included within this current EPIC study should therefore be considered as hypothesis generating, with more thorough disease subtype investigations being undertaken in future analyses.

4.3.3 Conclusion

In this comprehensive prospective analysis, higher predicted circulating 25(OH)D levels were associated with lower risks of: overall cancer incidence and mortality; colorectal cancer incidence; lung cancer incidence and mortality; kidney cancer incidence; stomach and oesophageal cancer incidence; pancreatic cancer incidence and mortality; thyroid cancer incidence; prostate cancer mortality; all-cause mortality; circulatory disease mortality; respiratory disease mortality; and digestive disease mortality. Many of these end-points have not previously been correlated with vitamin D levels, so these results provide indicators for future research on these diseases. However, care should be taken when interpreting these results due to the methodological limitations of observational research and of using predicted 25(OH)D scores.

4.4   Future research

Despite biological plausibility and a substantial amount of research, whether vitamin D reduces the incidence and mortality of certain chronic diseases - in particular cancer - remains uncertain. The predictor score results within this study provide new lines of inquiry for possible biological roles, in particular the inverse associations observed for thyroid cancer incidence, kidney cancer incidence, respiratory disease mortality and digestive disease mortality. Further observational research where predicted or actual 25(OH)D levels are measured is warranted to investigate these relationships. Additional prospective research is also required to scrutinise the elevated risks of all-cause mortality (33-35), overall cancer incidence (35;36), prostate cancer incidence (37-39), and pancreatic cancer incidence (40) observed at higher levels of actual circulating 25(OH)D which have previously been reported. These future analyses should measure actual circulating 25(OH)D levels, as predicted 25(OH)D scores have been shown to be poor at estimating absolute vitamin D levels.

The inverse associations observed for colorectal cancer incidence and circulatory disease mortality in this current analysis are consistent with the majority of previous prospective analyses. However, within an observational setting causality cannot be ascertained. A concern of measuring actual or predicted 25(OH)D levels is whether they are indicators of actual vitamin D levels or general "good health". Within EPIC, higher 25(OH)D levels were correlated with having a lower BMI, being younger and physically active, and reporting lower intakes of alcohol. Despite multivariable adjustments, residual confounding by these or other unmeasured factors may have caused the observed inverse associations.

Large RCTs are required to confirm these seemingly protective associations. In the U.S. one such intervention, called the VITamin D and OmegA-3 TriaL (VITAL) has just begun (202). For this double blind trial, 20,000 healthy men and women aged over 50 years have been recruited, half of whom will be receive 50 µg/day of vitamin D_3, with the remainder of participants receiving a placebo. The expected mean treatment period will be for 5 years and primary end-points considered are cancer and cardiovascular disease. Although RCTs are viewed as the highest form of evidence to prove a causal relationship, they are not without

limitations. Firstly, the previous WHI RCT - in which the intervention group were administered 10 μg of vitamin D plus 1 g of elemental calcium each day – was criticised for the 7 year trial duration being too short a time period for vitamin D to influence colorectal cancer occurrence (79;80), suggesting that the 5 year intervention within the VITAL study may be inadequate. Secondly, as nutritional supplements are freely available, there is a risk of individuals within the placebo group taking non-study supplements, and as a consequence the contrast in vitamin D levels with the intervention group may be too low to detect any effect. Finally, rarer outcomes cannot be captured with enough frequency to deduce any possible vitamin D effect.

An alternative approach, which may provide further indications of causal relationships, is to conduct suitably powered Mendelian randomisation analyses. Theodoratou *et al.,* (203) carried out one such analysis for colorectal cancer in a Scottish study containing 2,001 cases and 2,237 controls. In this analysis, as compared to homozygotes of the respective wild-type allele, carriers of a variant allele of the SNPs rs2282679 and rs12785878 were significantly associated with circulating 25(OH)D levels. The SNP rs2282679 is located in the gene which encodes a vitamin D binding protein that binds and transports vitamin D; while rs12785878 is located in a gene which encodes the enzyme 7-dehydrocholesterol reductase. Although neither of these SNPs (instrumental variables) was subsequently associated with colorectal cancer risk, this may have been the result of insufficient statistical power, which is a common pitfall of this study design (203). However, more recent evidence from a much larger pooled study of 10,061 cases and 12,768 controls from 13 studies also failed to observe relationships between these two SNPs and colorectal cancer risk (204). This analysis additionally investigated whether three other polymorphisms which have been associated with circulating 25(OH)D levels were related to colorectal cancer risk. These were the SNPs rs10741657 (located near the CYP2R1 gene), rs11234027 (located in a gene which encodes the enzyme 7-dehydrocholesterol reductase) and rs6013897 (located near the CYP24A1 gene). Similarly, none of these additional SNPs were associated with colorectal cancer risk. Within this pooled analysis just 5% of the variance in circulating 25(OH)D levels was explained by these studied SNPs (204). Thus, the possibility exists that even this large pooling study may have been underpowered to detect an association. An alternative interpretation is that the vitamin D-colorectal cancer relationship reported in observational studies is non-causal. Overall, this recent evidence highlights the importance of ongoing RCTs to provide insights into whether the vitamin D-cancer relationship is more cause-effect than correlation.

5. CONCLUSION

For the first time in a European population, predictor 25(OH)D scores were derived, validated and applied to the full EPIC cohort to assess the relationships with cancer incidence and mortality, circulatory disease mortality, respiratory disease mortality, digestive disease mortality and all-cause mortality. Significant inverse predicted 25(OH)D score associations were observed for: overall cancer incidence and mortality; colorectal cancer incidence; lung cancer incidence and mortality; kidney cancer incidence; stomach and oesophageal cancer incidence; pancreatic cancer incidence and mortality; thyroid cancer incidence; prostate cancer mortality; all-cause mortality; circulatory disease mortality; respiratory disease mortality; and digestive disease mortality. However, due to the methodological limitations specific to 25(OH)D predictor scores, and observational epidemiology in general, it is important to acknowledge that alternative explanations may explain some or all of these observed relationships. Nevertheless, the associations observed provide possible evidence for a beneficial role for vitamin D and the rarer outcomes have flagged up possible, previously unreported, relationships. Going forward, results from a recently begun VITAL RCT will hopefully provide insights into whether the vitamin D-cancer/circulatory disease relationships are causal, rather than mere correlations.

6. APPENDIX

Table A1. Risk (hazard ratios) of cancer incidence amongst men and women with and without adjustment for body mass index associated with a 25 nmol/L increment in predicted 25-hydroxyvitamin D level

	Both sexes	
	Multivariable Model	**Multivariable Model Minus BMI adjustment**
Men		
Women	Per 25 nmol/L increase	Per 25 nmol/L increase
Overall cancer incidence		
Both sexes		
Multivariable model - HR (95% CI) ‡	0.87 (0.80-0.92)	0.85 (0.80-0.91)
Multivariable model + physical activity adj - HR (95% CI) φ	0.92 (0.84-1.01)	0.89 (0.82-0.95)
Digestive system cancer incidence		
Both sexes		
Multivariable model - HR (95% CI) ‡	0.63 (0.52-0.76)	0.63 (0.54-0.73)
Multivariable model + physical activity adj - HR (95% CI) φ	0.61 (0.48-0.77)	0.62 (0.51-0.74)
Non-digestive system cancer incidence		
Both sexes		
Multivariable model - HR (95% CI) ‡	0.92 (0.85-0.99)	0.90 (0.84-0.96)
Multivariable model + physical activity adj - HR (95% CI) φ	0.99 (0.90-1.09)	0.94 (0.87-1.02)
Smoking related cancer incidence		
Both sexes		
Multivariable model - HR (95% CI) ‡	0.65 (0.56-0.74)	0.71 (0.64-0.80)
Multivariable model + physical activity adj - HR (95% CI) φ	0.69 (0.58-0.82)	0.77 (0.68-0.89)
Non-smoking related cancer incidence		
Both sexes		
Multivariable model - HR (95% CI) ‡	0.98 (0.89-1.07)	0.91 (0.84-0.98)
Multivariable model + physical activity adj - HR (95% CI) φ	1.03 (0.93-1.14)	0.92 (0.85-1.00)
Colorectal cancer		
Both sexes		
Multivariable model - HR (95% CI) † ¥	0.75 (0.57-0.97)	0.62 (0.50-0.78)
Multivariable model + physical activity adj - HR (95% CI) φ	0.83 (0.59-1.15)	0.63 (0.49-0.83)
Lung cancer		
Both sexes		
Multivariable model - HR (95% CI) †	0.64 (0.49-0.84)	1.01 (0.81-1.26)
Multivariable model + physical activity adj - HR (95% CI) φ	0.77 (0.55-1.08)	1.31 (1.02-1.68)
Kidney cancer		
Both sexes		
Multivariable model - HR (95% CI) †	0.47 (0.28-0.79)	0.42 (0.27-0.64)
Multivariable model + physical activity adj - HR (95% CI) φ	0.47 (0.24-0.92)	0.40 (0.24-0.67)
Stomach & oesophageal cancer		
Both sexes		
Multivariable model - HR (95% CI) †	0.60 (0.39-0.94)	0.74 (0.51-1.07)
Multivariable model + physical activity adj - HR (95% CI) φ	0.53 (0.30-0.94)	0.74 (0.48-1.16)
Bladder cancer		
Both sexes		
Multivariable model - HR (95% CI) †	0.89 (0.62-1.28)	0.83 (0.61-1.13)
Multivariable model + physical activity adj - HR (95% CI) φ	1.11 (0.70-1.75)	0.93 (0.63-1.36)
Pancreatic cancer		
Both sexes		
Multivariable model - HR (95% CI) †	0.52 (0.32-0.84)	0.51 (0.34-0.77)
Multivariable model + physical activity adj - HR (95% CI) φ	0.50 (0.27-0.94)	0.50 (0.31-0.81)
Liver cancer		
Both sexes		
Multivariable model - HR (95% CI) †	0.80 (0.41-1.43)	0.52 (0.31-0.87)
Multivariable model + physical activity adj - HR (95% CI) φ	0.69 (0.31-1.51)	0.42 (0.23-0.79)
Non-Hodgkin's lymphoma		
Both sexes		
Multivariable model - HR (95% CI) †	0.87 (0.54-1.43)	0.85 (0.57-1.28)
Multivariable model + physical activity adj - HR (95% CI) φ	0.81 (0.44-1.48)	0.80 (0.49-1.29)
Brain cancer		
Both sexes		
Multivariable model - HR (95% CI) †	0.77 (0.49-1.22)	0.72 (0.50-1.05)
Multivariable model + physical activity adj - HR (95% CI) φ	1.01 (0.59-1.72)	0.85 (0.55-1.32)
Skin cancer		
Both sexes		
Multivariable model - HR (95% CI) †	1.26 (1.00-1.60)	1.40 (1.16-1.69)
Multivariable model + physical activity adj - HR (95% CI) φ	1.20 (0.89-1.61)	1.39 (1.12-1.72)
Thyroid cancer		
Both sexes		
Multivariable model - HR (95% CI) †	0.50 (0.26-0.97)	0.69 (0.41-1.15)
Multivariable model + physical activity adj - HR (95% CI) φ	0.61 (0.27-1.36)	0.83 (0.46-1.51)
Prostate cancer		
Men		
Multivariable model - HR (95% CI) †	1.12 (0.86-1.47)	1.02 (0.80-1.30)
Multivariable model + physical activity adj - HR (95% CI) φ	1.30 (0.91-1.85)	1.08 (0.78-1.49)
Breast cancer		
Women		
N cases		
Multivariable model - HR (95% CI) † ¶	1.02 (0.89-1.17)	0.95 (0.85-1.06)
Multivariable model + physical activity adj - HR (95% CI) φ	1.12 (0.96-1.31)	0.99 (0.88-1.12)
Ovarian cancer		
Women		
N cases		
Multivariable model - HR (95% CI) † ¶	1.25 (0.85-1.85)	0.99 (0.71-1.36)
Multivariable model + physical activity adj - HR (95% CI) φ	1.22 (0.77-1.93)	0.93 (0.65-1.33)
Endometrial cancer		
Women		
N cases		
Multivariable model - HR (95% CI) † ¶	0.92 (0.63-1.35)	0.36 (0.27-0.49)
Multivariable model + physical activity adj - HR (95% CI) φ	0.83 (0.53-1.30)	0.28 (0.20-0.40)

† Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

¥ Multivariable model – plus adjustment for cereal fibre intake (g/day).

¶ Multivariable model – plus adjustment for age at first pregnancy (<21; 21-<30; 30+ years old; no children; or not specified); and age at menarche (<12; 12-<15; 15+ years old; or not specified).

ɸ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

BMI=body mass index.

Table A2. Risk (hazard ratios) of all-cause and cause-specific mortality amongst men and women with and without adjustment for body mass index associated with a 25 nmol/L increment in predicted 25-hydroxyvitamin D level

	Both sexes	
	Multivariable Model	Multivariable Model Minus BMI adjustment
Men		
Women	Per 25 nmol/L increase	Per 25 nmol/L increase
Total all-cause mortality		
Both sexes		
Multivariable model - HR (95% CI) †	0.48 (0.43-0.53)	0.52 (0.48-0.57)
Multivariable model + physical activity adj - HR (95% CI)	0.57 (0.50-0.64)	0.61 (0.56-0.67)
Cancer (C00-D48)		
Both sexes		
Multivariable model - HR (95% CI) †	0.65 (0.56-0.75)	0.72 (0.64-0.82)
Multivariable model + physical activity adj - HR (95% CI)	0.69 (0.57-0.83)	0.78 (0.68-0.91)
Circulatory diseases (I00-I99)		
Both sexes		
Multivariable model - HR (95% CI) †	0.39 (0.31-0.48)	0.30 (0.25-0.36)
Multivariable model + physical activity adj - HR (95% CI)	0.43 (0.33-0.58)	0.30 (0.24-0.38)
Respiratory diseases (J30-J98)		
Both sexes		
Multivariable model - HR (95% CI) †	0.12 (0.07-0.21)	0.38 (0.25-0.59)
Multivariable model + physical activity adj - HR (95% CI)	0.28 (0.14-0.55)	0.94 (0.57-1.56)
Digestive diseases (K00-K93)		
Both sexes		
Multivariable model - HR (95% CI) †	0.11 (0.06-0.20)	0.15 (0.09-0.24)
Multivariable model + physical activity adj - HR (95% CI)	0.19 (0.09-0.40)	0.24 (0.14-0.42)
External causes (S00-Y98)		
Both sexes		
Multivariable model - HR (95% CI) †	0.84 (0.53-1.33)	1.13 (0.77-1.65)
Multivariable model + physical activity adj - HR (95% CI)	0.75 (0.42-1.35)	1.17 (0.75-1.83)
Digestive system cancers *		
Both sexes		
Multivariable model - HR (95% CI) †	0.67 (0.49-0.91)	0.64 (0.49-0.82)
Multivariable model + physical activity adj - HR (95% CI)	0.70 (0.48-1.03)	0.65 (0.48-0.88)
Non-digestive system cancers *		
Both sexes		
Multivariable model - HR (95% CI) †	0.64 (0.54-0.76)	0.75 (0.65-0.87)
Multivariable model + physical activity adj - HR (95% CI)	0.68 (0.55-0.84)	0.83 (0.70-0.98)
Smoking related cancers *		
Both sexes		
Multivariable model - HR (95% CI) †	0.62 (0.51-0.77)	0.77 (0.65-0.91)
Multivariable model + physical activity adj - HR (95% CI)	0.67 (0.51-0.87)	0.86 (0.70-1.05)
Non-smoking related cancers *		
Both sexes		
Multivariable model - HR (95% CI) †	0.68 (0.55-0.85)	0.67 (0.56-0.80)
Multivariable model + physical activity adj - HR (95% CI)	0.72 (0.55-0.93)	0.70 (0.56-0.86)
Colorectal cancer (C18-C20)		
Both sexes		
Multivariable model - HR (95% CI) † ¥	0.76 (0.44-1.32)	0.69 (0.43-1.09)
Multivariable model + physical activity adj - HR (95% CI)	1.00 (0.52-1.94)	0.81 (0.47-1.39)
Pancreatic cancer (C25)		
Both sexes		
Multivariable model - HR (95% CI) †	0.58 (0.34-0.99)	0.51 (0.33-0.79)
Multivariable model + physical activity adj - HR (95% CI)	0.65 (0.33-1.25)	0.52 (0.31-0.88)
Lung cancer (C34)		
Both sexes		
Multivariable model - HR (95% CI) †	0.76 (0.55-1.05)	1.22 (0.94-1.58)
Multivariable model + physical activity adj - HR (95% CI)	0.91 (0.61-1.36)	1.55 (1.17-2.07)
Prostate cancer (C61)		
Both sexes		
Multivariable model - HR (95% CI) †	0.32 (0.12-0.84)	0.31 (0.13-0.73)
Multivariable model + physical activity adj - HR (95% CI)	0.25 (0.05-1.18)	0.26 (0.08-0.89)
Breast cancer (C50)		
Both sexes		
Multivariable model - HR (95% CI) † ¶	0.74 (0.46-1.19)	0.69 (0.41-1.09)
Multivariable model + physical activity adj - HR (95% CI)	0.95 (0.55-1.63)	0.77 (0.54-1.13)

† Multivariable model - Cox regression using body mass index (<22; 22-<25; 25-<30; 30-<35; 35+ kg/m^2), education status (none; primary school completed; technical/professional school; secondary school; longer education including university; or not specified), alcohol consumption (non-consumers; <5; 5-<15; 15-<30; 30+ g/day), smoking status and intensity (never; current , 1-15 cigarettes per day; current, 16-25 cigarettes per day; current, 16+ cigarettes per day; former, quit ≤10 years; former, quit 11-20 years; former, quit 20+ years; current, pipe/cigar/occasional; current/former, missing; unknown); smoking duration (<10; 10-<20; 20-<30; 30-<40; 40+ years; smoking duration unknown); ever use of contraceptive pill (yes; no; or unknown), menopausal status (premenopausal; postmenopausal; perimenopausal/unknown menopausal status; or surgical postmenopausal), and intakes of total energy (kcal/day), red and processed meat (g/day), fruits and vegetables (g/day), and dietary calcium (mg/day) (all continuous), and stratified by age (1-year categories), sex, and centre.

¥ Multivariable model – plus adjustment for cereal fibre intake (g/day).

¶ Multivariable model – plus adjustment for age at first pregnancy (<21; 21-<30; 30+ years old; no children; or not specified); and age at menarche (<12; 12-<15; 15+ years old; or not specified).

ɸ Multivariable model - plus adjustment for physical activity index (inactive; moderately inactive; moderately active; active).

BMI=body mass index.

7. REFERENCES

(1)   Garland CF, Garland FC. Do Sunlight and Vitamin D Reduce the Likelihood of Colon Cancer? International Journal of Epidemiology 1980 September 1;9(3):227-31.

(2)   Lund B, Sorenson OH, Bishop JE, Norman AW. Vitamin D Metabolism in Hypoparathyroidism. Journal of Clinical Endocrinology & Metabolism 1980 September 1;51(3):606-10.

(3)   Holick MF. Sunlight and vitamin D for bone health and prevention of autoimmune diseases, cancers, and cardiovascular disease. The American Journal of Clinical Nutrition 2004 December 1;80(6):1678S-88S.

(4)   Norman AW. From vitamin D to hormone D: fundamentals of the vitamin D endocrine system essential for good health. The American Journal of Clinical Nutrition 2008 August 1;88(2):491S-9S.

(5)   Haussler MR, Jurutka PW, Mizwicki M, Norman AW. Vitamin D receptor (VDR)-mediated actions of 1,25(OH)2vitamin D3: Genomic and non-genomic mechanisms. Best Practice & Research Clinical Endocrinology & Metabolism 2011 August;25(4):543-59.

(6)   Battault S, Whiting SJ, Peltier SL, Sadrin S, Gerber G, Maixent JM. Vitamin D metabolism, functions and needs: from science to health claims. Eur J Nutr 2013;52(2):429-41.

(7)   Ramagopalan SV, Heger A, Berlanga AJ, Maugeri NJ, Lincoln MR, Burrell A et al. A ChIP-seq defined genome-wide map of vitamin D receptor binding: Associations with disease and evolution. Genome Research 2010 October 1;20(10):1352-60.

(8)   Kostner K., Denzer N, Muller CS, Klein W, Reichrath J. The Relevance of Vitamin D Receptor (VDR) Gene Polymorphisms for Cancer: A Review of the Literature. Anticancer Research 2009 September 1;29(9):3511-36.

(9)   IARC. IARC - Vitamin D and Cancer. IARC Working Group Reports Vol.5. [5]. 2008.

(10)   McCullough ML, Bostick RM, Mayo TL. Vitamin D Gene Pathway Polymorphisms and Risk of Colorectal, Breast, and Prostate Cancer. Annu Rev Nutr 2009 July 22;29(1):111-32.

(11) Dusso AS, Brown AJ, Slatopolsky E. Vitamin D. American Journal of Physiology - Renal Physiology 2005 July 1;289(1):F8-F28.

(12) Hager G, Formanek M, Gedlicka C, Thurnher D, Knerer B, Kornfehl J. 1,25(OH)2 vitamin D3 induces elevated expression of the cell cycle-regulating genes P21 and P27 in squamous carcinoma cell lines of the head and neck. Acta Otolarymgol 2001;121(1):103-9.

(13) Fujioka T, Suzuki Y, Okamoto T, Mastushita N, Hasegawa M, Omori S. Prevention of renal cell carcinoma by active vitamin D3. World J Surg 2000;24(10):1205-10.

(14) Palmer HG, Gonzalez-Sancho J, Espada J, Berciano MT, Puig I, Baulida J et al. Vitamin D3 promotes the differentiation of colon carcinoma cells by the induction of E-cadherin and the inhibition of Beta-catenin signaling. The Journal of Cell Biology 2001 July 23;154(2):369-88.

(15) Deeb KK, Trump DL, Johnson CS. Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. Nat Rev Cancer 2007 September;7(9):684-700.

(16) Jiang F, Bao J, Li P, Nicosia SV, Bai W. Induction of Ovarian Cancer Cell Apoptosis by 1,25-Dihydroxyvitamin D3 through the Down-regulation of Telomerase. Journal of Biological Chemistry 2004 December 17;279(51):53213-21.

(17) Mantell DJ, Owens PE, Bundred NJ, Mawer EB, Canfield AE. 1,25-Dihydroxyvitamin D3 Inhibits Angiogenesis In Vitro and In Vivo. Circulation Research 2000 August 4;87(3):214-20.

(18) Xu H, Soruri A, Gieseler RK., Peters JH. 1,25-Dihydroxyvitamin D3 exerts opposing effects to IL-4 on MHC class-II antigen expression, accessory activity, and phagocytosis of human monocytes. Scand J Immunol 1993;38(6):535-40.

(19) Baeke F, Takiishi T, Korf H, Gysemans C, Mathieu C. Vitamin D: modulator of the immune system. Current Opinion in Pharmacology 2010 August;10(4):482-96.

(20) Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR et al. Toll-Like Receptor Triggering of a Vitamin D-Mediated Human Antimicrobial Response. Science 2006 March 24;311(5768):1770-3.

(21) Li YC, Kong J, Wei M, Chen ZF, Liu SQ, Cao LP. 1,25-Dihydroxyvitamin D3 is a negative endocrine regulator of the renin-angiotensin system. J Clin Invest 2002 July 15;110(2):229-38.

(22)   Zhao G, Simpson RU. Membrane localization, Caveolin-3 association and rapid
actions of vitamin D receptor in cardiac myocytes. Steroids 2010
August;75(8−9):555-9.

(23)   Maestro B, Molero S, Bajo S, Davila N, Calle C. Transcriptional activation of the
human insulin receptor gene by 1,25-dihydroxyvitamin D3. Cell Biochem Funct 2002
September 1;20(3):227-32.

(24)   Pittas AG, Lau J, Hu FB, Dawson-Hughes B. The Role of Vitamin D and Calcium in
Type 2 Diabetes. A Systematic Review and Meta-Analysis. Journal of Clinical
Endocrinology & Metabolism 2007 June 1;92(6):2017-29.

(25)   Olsen J.A., Loveridge N, Duthie GG, Shearer M.J. Fat-soluble vitamins, in Human
Nutrition and Dietetics, 10th edition. Garrow J.S, Ralph J.A., editors. 2000.

 (26)   Calvo MS, Whiting SJ, Barton CN. Vitamin D Intake: A Global Perspective of Current
Status. The Journal of Nutrition 2005 February 1;135(2):310-6.

(27)   Hypponen E, Power C. Hypovitaminosis D in British adults at age 45 y: nationwide
cohort study of dietary and lifestyle predictors. The American Journal of Clinical
Nutrition 2007 March 1;85(3):860-8.

(28)   Heaney RP, Davies KM, Chen TC, Holick MF, Barger-Lux MJ. Human serum 25-
hydroxycholecalciferol response to extended oral dosing with cholecalciferol. The
American Journal of Clinical Nutrition 2003 January 1;77(1):204-10.

(29)   Dawson-Hughes B, Heaney R, Holick M, Lips P, Meunier P, Vieth R. Estimates of
optimal vitamin D status. Osteoporos Int 2005;16(7):713-6.

(30)   SACN. SACN - Update on Vitamin D. Position statement by the Scientific Advisory
Committee on Nutrition, Science Advisory Committee on Nutrition. 2007. London,
TSO.

(31)   Kimball SM, Ursell MR, O'Connor P, Vieth R. Safety of vitamin D3 in adults with
multiple sclerosis. The American Journal of Clinical Nutrition 2007 September
1;86(3):645-51.

(32)   Vieth R. Vitamin D supplementation, 25-hydroxyvitamin D concentrations, and safety.
The American Journal of Clinical Nutrition 1999 May 1;69(5):842-56.

(33)   Durup D, Jorgensen HL, Christensen J, Schwarz P, Heegaard AM, Lind B. A
Reverse J-Shaped Association of All-Cause Mortality with Serum 25-Hydroxyvitamin
D in General Practice: The CopD Study. Journal of Clinical Endocrinology &
Metabolism 2012 August 1;97(8):2644-52.

(34)   Melamed ML, Michos ED, Post W, Astor B. 25-hydroxyvitamin d levels and the risk of mortality in the general population. Archives of Internal Medicine 2008 August 11;168(15):1629-37.

(35)   Michaelsson K, Baron JA, Snellman G, Gedeborg R, Byberg L, Sundstrom J et al. Plasma vitamin D and mortality in older men: a community-based prospective cohort study. The American Journal of Clinical Nutrition 2010 October 1;92(4):841-8.

(36)   Freedman DM, Looker AC, Abnet CC, Linet MS, Graubard BI. Serum 25-Hydroxyvitamin D and Cancer Mortality in the NHANES III Study (1988−2006). Cancer Research 2010 November 1;70(21):8587-97.

(37)   Tuohimaa P, Tenkanen L, Ahonen M, Lumme S, Jellum E, Hallmans G+ et al. Both high and low levels of blood vitamin D are associated with a higher prostate cancer risk: A longitudinal, nested case-control study in the Nordic countries. Int J Cancer 2004 January 1;108(1):104-8.

(38)   Albanes D, Mondul AM, Yu K, Parisi D, Horst RL, Virtamo J et al. Serum 25-Hydroxy Vitamin D and Prostate Cancer Risk in a Large Nested Case−Control Study. Cancer Epidemiology Biomarkers & Prevention 2011 September 1;20(9):1850-60.

(39)   Meyer HE, Robsahm TE, Bjorge T, Brustad M, Blomhoff R. Vitamin D, season, and risk of prostate cancer: a nested case-control study within Norwegian health studies. The American Journal of Clinical Nutrition 2013 January 1;97(1):147-54.

(40)   Stolzenberg-Solomon RZ, Jacobs EJ, Arslan AA, Qi D, Patel AV, Helzlsouer KJ et al. Circulating 25-Hydroxyvitamin D and Risk of Pancreatic Cancer: Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. American Journal of Epidemiology 2010 July 1;172(1):81-93.

(41)   Doets E, de Wit L, Dhonukshe-Rutten R, Cavelaars A, Raats M, Timotijevic L et al. Current micronutrient recommendations in Europe: towards understanding their differences and similarities. Eur J Nutr 2008;47(1):17-40.

(42)   FAO/WHO. FAO/WHO Expert Consultation on Human Vitamin and Mineral Requirements. 1998.

(43)   Webb AR, Holick MF. The Role of Sunlight in the Cutaneous Production of Vitamin D3. Annu Rev Nutr 1988 July 1;8(1):375-99.

(44)   Carnevale V, Modoni S, Pileri M, Di Giorgio A, Chiodini I, Minisola S et al. Longitudinal Evaluation of Vitamin D Status in Healthy Subjects from Southern Italy: Seasonal and Gender Differences. Osteoporos Int 2001;12(12):1026-30.

(45)   Vik TF, Try KF, Stromme JH. The vitamin D status of man at 70 degrees north. Scand J Clin Invest 1980 May 1;40(3):227-32.

(46)   Chapuy MC, Preziosi P, Maamer M, Arnaud S, Galan P, Hercberg S et al. Prevalence of Vitamin D Insufficiency in an Adult Normal Population. Osteoporos Int 1997;7(5):439-43.

(47)   Mullie P, Gandini S, Boniol M, Autier P. Latitude of residence in Europe and serum 25-hydroxy vitamin D levels: a systematic review. 201-210. 2008. Vitamin D and Cancer. IARC Working Group Reports, vol. 5. WHO-International Agency for Research on Cancer.

(48)   Hatun S, Islam O, Cizmecioglu F, Kara B, Babaoglu K, Berk F et al. Subclinical Vitamin D Deficiency Is Increased in Adolescent Girls Who Wear Concealing Clothing. The Journal of Nutrition 2005 February 1;135(2):218-22.

(49)   McCullough ML, Weinstein SJ, Freedman DM, Helzlsouer K, Flanders WD, Koenig K et al. Correlates of Circulating 25-Hydroxyvitamin D: Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. American Journal of Epidemiology 2010 July 1;172(1):21-35.

(50)   Millen AE, Wactawski-Wende J, Pettinger M, Melamed ML, Tylavsky FA, Liu S et al. Predictors of serum 25-hydroxyvitamin D concentrations among postmenopausal women: the Women's Health Initiative Calcium plus Vitamin D Clinical Trial. The American Journal of Clinical Nutrition 2010 May 1;91(5):1324-35.

(51)   Wortsman J, Matsuoka LY, Chen TC, Lu Z, Holick MF. Decreased bioavailability of vitamin D in obesity. The American Journal of Clinical Nutrition 2000 September 1;72(3):690-3.

(52)   Giovannucci E, Liu Y, Rimm EB, Hollis BW, Fuchs CS, Stampfer MJ et al. Prospective Study of Predictors of Vitamin D Status and Cancer Incidence and Mortality in Men. Journal of the National Cancer Institute 2006 April 5;98(7):451-9.

(53)   Scragg R, Sowers M, Bell C. Serum 25-Hydroxyvitamin D, Diabetes, and Ethnicity in the Third National Health and Nutrition Examination Survey. Diabetes Care 2004 December 1;27(12):2813-8.

(54)   Cawthon PM, Parimi N, Barrett-Connor E, Laughlin GA, Ensrud KE, Hoffman AR et al. Serum 25-Hydroxyvitamin D, Parathyroid Hormone, and Mortality in Older Men. Journal of Clinical Endocrinology & Metabolism 2010 October 1;95(10):4625-34.

(55)   Hermann AP, Brot C, Gram J, Kolthoff N, Mosekilde L. Premenopausal Smoking and Bone Density in 2015 Perimenopausal Women. J Bone Miner Res 2000 April 1;15(4):780-7.

(56)   Giovannucci E. The epidemiology of vitamin D and cancer incidence and mortality: A review (United States). Cancer Causes Control 2005;16(2):83-95.

(57)   Ordonez-Mena J, Schottker B, Haug U, Muller H, Kohrle J, Schomburg L et al. Serum 25-Hydroxyvitamin D and Cancer Risk in Older Adults: Results from a Large German Prospective Cohort Study. Cancer Epidemiology Biomarkers & Prevention 2013 May 1;22(5):905-16.

(58)   Lappe JM, Travers-Gustafson D, Davies KM, Recker RR, Heaney RP. Vitamin D and calcium supplementation reduces cancer risk: results of a randomized trial. The American Journal of Clinical Nutrition 2007 June 1;85(6):1586-91.

(59)   Schottker B, Haug U, Schomburg L, Kohrle J, Perna L, Muller H et al. Strong associations of 25-hydroxyvitamin D concentrations with all-cause, cardiovascular, cancer, and respiratory disease mortality in a large cohort study. The American Journal of Clinical Nutrition 2013 April 1;97(4):782-93.

(60)   Pilz S, Dobnig H, Winklhofer-Roob B, Riedmuller G, Fischer JE, Seelhorst U et al. Low Serum Levels of 25-Hydroxyvitamin D Predict Fatal Cancer in Patients Referred to Coronary Angiography. Cancer Epidemiology Biomarkers & Prevention 2008 May 1;17(5):1228-33.

(61)   Tomson J, Emberson J, Hill M, Gordon A, Armitage J, Shipley M et al. Vitamin D and risk of death from vascular and non-vascular causes in the Whitehall study and meta-analyses of 12 000 deaths. European Heart Journal 2013 May 7;34(18):1365-74.

(62)   Skaaby T, Husemoen LLN, Pisinger C, Jorgensen T, Thuesen BH, Fenger M et al. Vitamin D Status and Cause-Specific Mortality: A General Population Study. PLoS ONE 2012 December 20;7(12):e52423.

(63)   Hutchinson MS, Grimnes G, Joakimsen RM, Figenschau Y, Jorde R. Low serum 25-hydroxyvitamin D levels are associated with increased all-cause mortality risk in a general population: the Tromso study. European Journal of Endocrinology 2010 May 1;162(5):935-42.

(64)   Signorello LB, Han X, Cai Q, Cohen SS, Cope EL, Zheng W et al. A Prospective Study of Serum 25-Hydroxyvitamin D Levels and Mortality Among African Americans and Non-African Americans. American Journal of Epidemiology 2013 January 15;177(2):171-9.

(65)   Eaton CB, Young A, Allison MA, Robinson J, Martin LW, Kuller LH et al. Prospective
         association of vitamin D concentrations with mortality in postmenopausal women:
         results from the Women's Health Initiative (WHI). The American Journal of Clinical
         Nutrition 2011 December 1;94(6):1471-8.

(66)   Kesse E, Boutron-Ruault MC, Norat T, Riboli E, Clavel-Chapelon F, Group N. Dietary
         calcium, phosphorus, vitamin D, dairy products and the risk of colorectal adenoma
         and cancer among French women of the E3N-EPIC prospective study. Int J Cancer
         2005 October 20;117(1):137-44.

(67)   Jenab M, Bueno-de-Mesquita HB, Ferrari P, van Duijnhoven FJB, Norat T, Pischon T
         et al. Association between pre-diagnostic circulating vitamin D concentration and risk
         of colorectal cancer in European populations:a nested case-control study. BMJ 2010
         January 21;340.

(68)   McCullough ML, Robertson AS, Rodriguez C, Jacobs EJ, Chao A, Carolyn J et al.
         Calcium, vitamin D, dairy products, and risk of colorectal cancer in the Cancer
         Prevention Study II Nutrition Cohort (United States). Cancer Causes Control
         2003;14(1):1-12.

(69)   Park SY, Murphy SP, Wilkens LR, Nomura AMY, Henderson BE, Kolonel LN.
         Calcium and Vitamin D Intake and Risk of Colorectal Cancer: The Multiethnic Cohort
         Study. American Journal of Epidemiology 2007 April 1;165(7):784-93.

(70)   Wu K, Feskanich D, Fuchs CS, Willett WC, Hollis BW, Giovannucci EL. A Nested
         Case−Control Study of Plasma 25-Hydroxyvitamin D Concentrations and Risk of
         Colorectal Cancer. Journal of the National Cancer Institute 2007 July
         18;99(14):1120-9.

(71)   Woolcott CG, Wilkens LR, Nomura AMY, Horst RL, Goodman MT, Murphy SP et al.
         Plasma 25-Hydroxyvitamin D Levels and the Risk of Colorectal Cancer: The
         Multiethnic Cohort Study. Cancer Epidemiology Biomarkers & Prevention 2010
         January 1;19(1):130-4.

(72)   Garland CF., Comstock GW., Garland FC., Helsing KJ., Shaw EK., Gorham ED.
         Serum 25-hydroxyvitamin D and colon cancer: eight-year prospective study. Lancet
         1989;2(8673):1176-8.

(73)   Braun MM, Helzlsouer KJ, Hollis BW, Comstock GW. Colon Cancer and Serum
         Vitamin D Metabolite Levels 10-17 Years prior to Diagnosis. American Journal of
         Epidemiology 1995 September 15;142(6):608.

(74) Lee JE, Li H, Chan AT, Hollis BW, Lee IM, Stampfer MJ et al. Circulating Levels of Vitamin D and Colon and Rectal Cancer: The Physicians' Health Study and a Meta-analysis of Prospective Studies. Cancer Prevention Research 2011 May 1;4(5):735-43.

(75) Weinstein SJ, Yu K, Horst RL, Ashby J, Virtamo J, Albanes D. Serum 25-Hydroxyvitamin D and Risks of Colon and Rectal Cancer in Finnish Men. American Journal of Epidemiology 2011 March 1;173(5):499-508.

(76) Otani T, Iwasaki M, Sasazuki S, Inoue M, Tsugane S. Plasma vitamin D and risk of colorectal cancer: the Japan Public Health Center-Based Prospective Study. Br J Cancer 2007 July 10;97(3):446-51.

(77) Touvier M, Chan DSM, Lau R, Aune D, Vieira R, Greenwood DC et al. Meta-Analyses of Vitamin D Intake, 25-Hydroxyvitamin D Status, Vitamin D Receptor Polymorphisms, and Colorectal Cancer Risk. Cancer Epidemiology Biomarkers & Prevention 2011 May 1;20(5):1003-16.

(78) Trivedi DP, Doll R, Khaw KT. Effect of four monthly oral vitamin D3 (cholecalciferol) supplementation on fractures and mortality in men and women living in the community: randomised double blind controlled trial. BMJ 2003 March 1;326.

(79) Wactawski-Wende J, Kotchen JM, Anderson GL, Assaf AR, Brunner RL, O'Sullivan MJ et al. Calcium plus Vitamin D Supplementation and the Risk of Colorectal Cancer. N Engl J Med 2006 February 16;354(7):684-96.

(80) Giovannucci E. Epidemiology of vitamin D and colorectal cancer: Casual or causal link? The Journal of Steroid Biochemistry and Molecular Biology 2010 July;121(1−2):349-54.

(81) Boscoe F, Schymura M. Solar ultraviolet-B exposure and cancer incidence and mortality in the United States, 1993-2002. BMC Cancer 2006;6(1):264.

(82) Ng K, Meyerhardt JA, Wu K, Feskanich D, Hollis BW, Giovannucci EL et al. Circulating 25-Hydroxyvitamin D Levels and Survival in Patients With Colorectal Cancer. Journal of Clinical Oncology 2008 June 20;26(18):2984-91.

(83) Fedirko V, Riboli E, Tjonneland A, Ferrari P, Olsen A, Bueno-de-Mesquita HB et al. Prediagnostic 25-Hydroxyvitamin D, VDR and CASR Polymorphisms, and Survival in Patients with Colorectal Cancer in Western European Populations. Cancer Epidemiology Biomarkers & Prevention 2012 April 1;21(4):582-93.

(84) Mezawa H, Sugiura T, Watanabe M, Norizoe C, Takahashi D, Shimojima A et al. Serum vitamin D levels and survival of patients with colorectal cancer: Post-hoc analysis of a prospective cohort study. BMC Cancer 2010;10(1):347.

(85) Ng K, Wolpin BM, Meyerhardt JA, Wu K, Chan AT, Hollis BW et al. Prospective study of predictors of vitamin D status and survival in patients with colorectal cancer. Br J Cancer 2009 August 18;101(6):916-23.

(86) Schwartz G, Hanchette C. UV, latitude, and spatial trends in prostate cancer mortality: All sunlight is not the same (United States). Cancer Causes Control 2006;17(8):1091-101.

(87) Chan JM FAU, Pietinen PF, Virtanen MF, Malila NF, Tangrea JF, Albanes DF et al. Diet and prostate cancer risk in a cohort of smokers, with a specific focus on calcium and phosphorus (Finland). Cancer Causes Control 2000;11(9):859-67.

(88) Chan JM FAU, Giovannucci E FAU - Andersson, Andersson SO FAU, Yuen JF, Adami HO FAU, Wolk A. Dairy products, calcium, phosphorous, vitamin D, and risk of prostate cancer (Sweden). Cancer Causes Control 1998;9(6):559-66.

(89) Kristal AR, Cohen JH, Qu P, Stanford JL. Associations of Energy, Fat, Calcium, and Vitamin D with Prostate Cancer Risk. Cancer Epidemiology Biomarkers & Prevention 2002 August 1;11(8):719-25.

(90) Braun MM FAU, Helzlsouer KJ FAU, Hollis BW FAU, Comstock GW. Prostate cancer and prediagnostic levels of serum vitamin D metabolites (Maryland, United States). Cancer Causes Control 1995;6(3):235-9.

(91) Nomura AM., Stemmermann GN., Lee J, Kolonel LN.., Chen TC., Turner A et al. Serum vitamin D)metabolite levels and the subsequent development of prostate cancer (Hawaii, United States. Cancer Causes Control 1998;9(4):425-32.

(92) Platz EA, Leitzmann MF, Hollis BW, Willett WC, Giovannucci E. Plasma 1,25-dihydroxy- and 25-hydroxyvitamin D and subsequent risk of prostate cancer. Cancer Causes Control 2004;15(3):255-65.

(93) Jacobs ET, Giuliano AR, Martinez M, Hollis BW, Reid ME, Marshall JR. Plasma levels of 25-hydroxyvitamin D, 1,25-dihydroxyvitamin D and the risk of prostate cancer. The Journal of Steroid Biochemistry and Molecular Biology 2004 May;89−90(0):533-7.

(94)   Baron JA, Beach M, Wallace K, Grau MV, Sandler RS, Mandel JS et al. Risk of Prostate Cancer in a Randomized Clinical Trial of Calcium Supplementation. Cancer Epidemiology Biomarkers & Prevention 2005 March 1;14(3):586-9.

(95)   Ahn J, Peters U, Albanes D, Purdue MP, Abnet CC, Chatterjee N et al. Serum Vitamin D Concentration and Prostate Cancer Risk: A Nested Case−Control Study. Journal of the National Cancer Institute 2008 June 4;100(11):796-804.

(96)   Faupel-Badger JM, Diaw L, Albanes D, Virtamo J, Woodson K, Tangrea JA. Lack of Association between Serum Levels of 25-Hydroxyvitamin D and the Subsequent Risk of Prostate Cancer in Finnish Men. Cancer Epidemiology Biomarkers & Prevention 2007 December 1;16(12):2784-6.

(97)   Travis RC, Crowe FL, Allen NE, Appleby PN, Roddam AW, Tjonneland A et al. Serum Vitamin D and Risk of Prostate Cancer in a Case-Control Analysis Nested Within the European Prospective Investigation into Cancer and Nutrition (EPIC). American Journal of Epidemiology 2009 May 15;169(10):1223-32.

(98)   Barnett CM, Nielson CM, Shannon J, Chan JM, Shikany JM, Bauer DC et al. Serum 25-OH vitamin D levels and risk of developing prostate cancer in older men. Cancer Causes Control 2010;21(8):1297-303.

(99)   Gilbert R, Metcalfe C, Fraser WD, Donovan J, Hamdy F, Neal DE et al. Associations of circulating 25-hydroxyvitamin D with prostate cancer diagnosis, stage and grade. Int J Cancer 2012 September 1;131(5):1187-96.

(100) Gilbert R, Martin R, Beynon R, Harris R, Savovic J, Zuccolo L et al. Associations of circulating and dietary vitamin D with prostate cancer risk: a systematic review and dose−response meta-analysis. Cancer Causes Control 2011;22(3):319-40.

(101) Tretli S, Hernes E, Berg JP, Hestvik UE, Robsahm TE. Association between serum 25(OH)D and death from prostate cancer. Br J Cancer 2009 January 20;100(3):450-4.

(102) Shui IM, Mucci LA, Kraft P, Tamimi RM, Lindstrom S, Penney KL et al. Vitamin D−Related Genetic Variation, Plasma Vitamin D, and Risk of Lethal Prostate Cancer: A Prospective Nested Case−Control Study. Journal of the National Cancer Institute 2012 May 2;104(9):690-9.

(103) Gorham ED, Garland FC, Garland CF. Sunlight and breast cancer incidence in the USSR. Int J Epidemiol 1990;19:820-4.

(104)   Grant WB. An ecologic study of dietary and solar ultraviolet-B links to breast carcinoma mortality rates. Cancer 2002 January 1;94(1):272-81.

(105)   Colston KW., Lowe LC., Mansi JL., Campbell MJ. Vitamin D Status and Breast Cancer Risk. Anticancer Research 2006 July 1;26(4A):2573-80.

(106)   Palmieri C, MacGregor T, Girgis S, Vigushin D. Serum 25-hydroxyvitamin D levels in early and advanced breast cancer. Journal of Clinical Pathology 2006 December 1;59(12):1334-6.

(107)   Abbas S, Chang-Claude J, Linseisen J. Plasma 25-hydroxyvitamin D and premenopausal breast cancer risk in a German case-control study. Int J Cancer 2009;124:250-5.

(108)   Abbas S, Linseisen J, Slanger T, Kropp S, Mutschelknauss EJ, Flesch-Janys D et al. Serum 25-hydroxyvitamin D and risk of post-menopausal breast cancer--results of a large case-control study. Carcinogenesis 2008;29:93-9.

(109)   Crew KD, Gammon MD, Steck SE, Hershman DL, Cremers S, Dworakowski E et al. Association between Plasma 25-Hydroxyvitamin D and Breast Cancer Risk. Cancer Prevention Research 2009 June 1;2(6):598-604.

(110)   Rejnmark L, Tietze A, Vestergaard P, Buhl L, Lehbrink M, Heickendorff L et al. Reduced Prediagnostic 25-Hydroxyvitamin D Levels in Women with Breast Cancer: A Nested Case-Control Study. Cancer Epidemiology Biomarkers & Prevention 2009 October 1;18(10):2655-60.

(111)   Engel P, Fagherazzi G, Boutten A, Dupre T, Mesrine S, Boutron-Ruault MC et al. Serum 25(OH) Vitamin D and Risk of Breast Cancer: A Nested Case-Control Study from the French E3N Cohort. Cancer Epidemiology Biomarkers & Prevention 2010 September 1;19(9):2341-50.

(112)   Bertone-Johnson ER, Chen WY, Holick MF, Hollis BW, Colditz GA, Willett WC et al. Plasma 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2005;14:1991-7.

(113)   Freedman DM, Chang SC, Falk RT, Purdue MP, Huang WY, McCarty CA et al. Serum Levels of Vitamin D Metabolites and Breast Cancer Risk in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. Cancer Epidemiology Biomarkers & Prevention 2008 April 1;17(4):889-94.

(114)   Chlebowski RT, Johnson KC, Kooperberg C, Pettinger M, Wactawski-Wende J, Rohan T et al. Calcium Plus Vitamin D Supplementation and the Risk of Breast

Cancer. Journal of the National Cancer Institute 2008 November 19;100(22):1581-91.

(115) McCullough M, Stevens V, Patel R, Jacobs E, Bain E, Horst R et al. Serum 25-hydroxyvitamin D concentrations and postmenopausal breast cancer risk: a nested case control study in the Cancer Prevention Study-II Nutrition Cohort. Breast Cancer Research 2009;11(4):R64.

(116) Eliassen AH, Spiegelman D, Hollis B, Horst R, Willett W, Hankinson S. Plasma 25-hydroxyvitamin D and risk of breast cancer in the Nurses' Health Study II. Breast Cancer Research 2011;13(3):R50.

(117) Kuhn T, Kaaks R, Becker S, Eomois PP, Clavel-Chapelon F, Kvaskoff M et al. Plasma 25-hydroxyvitamin D and the risk of breast cancer in the European prospective investigation into cancer and nutrition: A nested case–control study. Int J Cancer 2013 October 1;133(7):1689-700.

(118) Mohr SB, Gorham ED, Alcaraz JE, Kane CJ, Macera CA, Parsons J et al. Serum 25-Hydroxyvitamin D and Prevention of Breast Cancer: Pooled Analysis. Anticancer Research 2011 September 1;31(9):2939-48.

(119) Goodwin PJ, Ennis M, Pritchard KI, Koo J, Hood N. Prognostic effects of 25-hydroxyvitamin D levels in early breast cancer. J Clin Oncol 2009;27:3757-63.

(120) Vrieling A, Hein R, Abbas S, Schneeweiss A, Flesch-Janys D, Chang-Claude J. Serum 25-hydroxyvitamin D and postmenopausal breast cancer survival: a prospective patient cohort study. Breast Cancer Research 2011;13(4):R74.

(121) Jacobs ET, Thomson CA, Flatt SW, Al-Delaimy WK, Hibler EA, Jones LA et al. Vitamin D and breast cancer recurrence in the Women's Healthy Eating and Living (WHEL) Study. Am J Clin Nutr 2011;93:108-17.

(122) Stolzenberg-Solomon RZ, Vieth R, Azad A, Pietinen P, Taylor PR, Virtamo J et al. A Prospective Nested Case-Control Study of Vitamin D Status and Pancreatic Cancer Risk in Male Smokers. Cancer Research 2006 October 15;66(20):10213-9.

(123) Stolzenberg-Solomon RZ, Hayes RB, Horst RL, Anderson KE, Hollis BW, Silverman DT. Serum Vitamin D and Risk of Pancreatic Cancer in the Prostate, Lung, Colorectal, and Ovarian Screening Trial. Cancer Research 2009 February 15;69(4):1439-47.

(124)   Bao Y, Ng K, Wolpin BM, Michaud DS, Giovannucci E, Fuchs CS. Predicted vitamin
        D status and pancreatic cancer risk in two prospective cohort studies. Br J Cancer
        2010 April 27;102(9):1422-7.

(125)   Wolpin BM, Ng K, Bao Y, Kraft P, Stampfer MJ, Michaud DS et al. Plasma 25-
        Hydroxyvitamin D and Risk of Pancreatic Cancer. Cancer Epidemiology Biomarkers
        & Prevention 2012 January 1;21(1):82-91.

(126)   Kilkkinen A, Knekt P, Heliovaara M, Rissanen H, Marniemi J, Hakulinen T et al.
        Vitamin D Status and the Risk of Lung Cancer: A Cohort Study in Finland. Cancer
        Epidemiology Biomarkers & Prevention 2008 November 1;17(11):3274-8.

(127)   Weinstein SJ, Yu K, Horst RL, Parisi D, Virtamo J, Albanes D. Serum 25-
        Hydroxyvitamin D and Risk of Lung Cancer in Male Smokers: A Nested Case-Control
        Study. PLoS ONE 2011 June 10;6(6):e20796.

(128)   Cheng TY, Neuhouser M. Serum 25-hydroxyvitamin D, vitamin A, and lung cancer
        mortality in the US population: a potential nutrient–nutrient interaction. Cancer
        Causes Control 2012;23(9):1557-65.

(129)   Gallicchio L, Moore LE, Stevens VL, Ahn J, Albanes D, Hartmuller V et al. Circulating
        25-Hydroxyvitamin D and Risk of Kidney Cancer: Cohort Consortium Vitamin D
        Pooling Project of Rarer Cancers. American Journal of Epidemiology 2010 July
        1;172(1):47-57.

(130)   Joh H, Giovannucci EL, Bertrand KA, Lim S, Cho E. Predicted Plasma 25-
        Hydroxyvitamin D and Risk of Renal Cell Cancer. Journal of the National Cancer
        Institute 2013 May 15;105(10):726-32.

(131)   Abnet CC, Chen Y, Chow WH, Gao YT, Helzlsouer KJ, Le Marchand Lc et al.
        Circulating 25-Hydroxyvitamin D and Risk of Esophageal and Gastric Cancer: Cohort
        Consortium Vitamin D Pooling Project of Rarer Cancers. American Journal of
        Epidemiology 2010 July 1;172(1):94-106.

(132)   Chen W, Dawsey SM, Qiao YL, Mark SD, Dong ZW, Taylor PR et al. Prospective
        study of serum 25(OH)-vitamin D concentration and risk of oesophageal and gastric
        cancers. Br J Cancer 2007 June 5;97(1):123-8.

(133)   Mondul AM, Weinstein SJ, Mannisto S, Snyder K, Horst RL, Virtamo J et al. Serum
        Vitamin D and Risk of Bladder Cancer. Cancer Research 2010 November
        15;70(22):9218-23.

(134) Mondul AM, Weinstein SJ, Horst RL, Purdue M, Albanes D. Serum Vitamin D and Risk of Bladder Cancer in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Cancer Epidemiology Biomarkers & Prevention 2012 July 1;21(7):1222-5.

(135) Purdue MP, Freedman DM, Gapstur SM, Helzlsouer KJ, Laden F, Lim U et al. Circulating 25-Hydroxyvitamin D and Risk of Non-Hodgkin Lymphoma: Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. American Journal of Epidemiology 2010 July 1;172(1):58-69.

(136) Luczynska A, Kaaks R, Rohrmann S, Becker S, Linseisen J, Buijsse B et al. Plasma 25-hydroxyvitamin D concentration and lymphoma risk: results of the European Prospective Investigation into Cancer and Nutrition. The American Journal of Clinical Nutrition 2013 September 1.

(137) Major JM, Kiruthu C, Weinstein SJ, Horst RL, Snyder K, Virtamo J et al. Pre-Diagnostic Circulating Vitamin D and Risk of Melanoma in Men. PLoS ONE 2012 April 25;7(4):e35112.

(138) van der Pols JC, Russell A, Bauer U, Neale RE, Kimlin MG, Green AC. Vitamin D Status and Skin Cancer Risk Independent of Time Outdoors: 11-Year Prospective Study in an Australian Community. J Invest Dermatol 2013 March;133(3):637-41.

(139) Tang J, Parimi N, Wu A, John Boscardin W, Shikany J, Chren MM et al. Inverse association between serum 25(OH) vitamin D levels and non-melanoma skin cancer in elderly men. Cancer Causes Control 2010;21(3):387-91.

(140) Liang G, Nan H, Qureshi AA, Han J. Pre-Diagnostic Plasma 25-Hydroxyvitamin D Levels and Risk of Non-Melanoma Skin Cancer in Women. PLoS ONE 2012 April 6;7(4):e35211.

(141) Tworoger SS, Lee IM, Buring JE, Rosner B, Hollis BW, Hankinson SE. Plasma 25-Hydroxyvitamin D and 1,25-Dihydroxyvitamin D and Risk of Incident Ovarian Cancer. Cancer Epidemiology Biomarkers & Prevention 2007 April 1;16(4):783-8.

(142) Arslan AA, Clendenen TV, Koeniq KL, Hultdin J, Enquist K, Agren A et al. Circulating vitamin d and risk of epithelial ovarian cancer. J Oncol 2009;Epub(672492).

(143) Zheng W, Danforth KN, Tworoger SS, Goodman MT, Arslan AA, Patel AV et al. Circulating 25-Hydroxyvitamin D and Risk of Epithelial Ovarian Cancer: Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. American Journal of Epidemiology 2010 July 1;172(1):70-80.

(144)   Yin L, Grandi N, Raum E, Haug U, Arndt V, Brenner H. Meta-analysis: Circulating vitamin D and ovarian cancer risk. Gynecologic Oncology 2011 May 1;121(2):369-75.

(145)   Zeleniuch-Jacquotte A, Gallicchio L, Hartmuller V, Helzlsouer KJ, McCullough ML, Setiawan VW et al. Circulating 25-Hydroxyvitamin D and Risk of Endometrial Cancer: Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. American Journal of Epidemiology 2010 July 1;172(1):36-46.

(146)   Liu JJ, Bertrand KA, Karageorgi S, Giovannucci E, Hankinson SE, Rosner B et al. Prospective analysis of vitamin D and endometrial cancer risk. Annals of Oncology 2013 March 1;24(3):687-92.

(147)   Zittermann A, Schleithoff SS, Frisch S, Gotting C, Kuhn J, Koertke H et al. Circulating Calcitriol Concentrations and Total Mortality. Clinical Chemistry 2009 June 1;55(6):1163-70.

(148)   Dobnig H, Pilz S, Scharnagl H, Renner W, Seelhorst U, Wellnitz B et al. Independent association of low serum 25-hydroxyvitamin d and 1,25-dihydroxyvitamin d levels with all-cause and cardiovascular mortality. Archives of Internal Medicine 2008 June 23;168(12):1340-9.

(149)   Virtanen J, Nurmi T, Voutilainen S, Mursu J, Tuomainen TP. Association of serum 25-hydroxyvitamin D with the risk of death in a general older population in Finland. Eur J Nutr 2011;50(5):305-12.

(150)   Pilz S, Dobnig H, Nijpels G, Heine RJ, Stehouwer CDA, Snijder MB et al. Vitamin D and mortality in older men and women. Clinical Endocrinology 2009 November 1;71(5):666-72.

(151)   Ginde AA, Scragg R, Schwartz RS, Camargo CA. Prospective Study of Serum 25-Hydroxyvitamin D Level, Cardiovascular Disease Mortality, and All-Cause Mortality in Older U.S. Adults. Journal of the American Geriatrics Society 2009 September 1;57(9):1595-603.

(152)   Visser M, Deeg DJ, Puts MT, Seidell JC, Lips P. Low serum concentrations of 25-hydroxyvitamin D in older persons and the risk of nursing home admission. The American Journal of Clinical Nutrition 2006 September 1;84(3):616-22.

(153)   Autier P, Gandini S. Vitamin D supplementation and total mortality: A meta-analysis of randomized controlled trials. Archives of Internal Medicine 2007 September 10;167(16):1730-7.

(154) Kilkkinen A, Knekt P, Aro A, Rissanen H, Marniemi J, Heliovaara M et al. Vitamin D Status and the Risk of Cardiovascular Disease Death. American Journal of Epidemiology 2009 October 15;170(8):1032-9.

(155) Pilz S, Marz W, Wellnitz B, Seelhorst U, Fahrleitner-Pammer A, Dimai HP et al. Association of Vitamin D Deficiency with Heart Failure and Sudden Cardiac Death in a Large Cross-Sectional Study of Patients Referred for Coronary Angiography. Journal of Clinical Endocrinology & Metabolism 2008 October 1;93(10):3927-35.

(156) Semba RD, Houston DK, Bandinelli S, Sun K, Cherubini A, Cappola AR et al. Relationship of 25-hydroxyvitamin D with all-cause and cardiovascular disease mortality in older community-dwelling adults. Eur J Clin Nutr 2009 December 2;64(2):203-9.

(157) Bolland MJ, Bacon CJ, Horne AM, Mason BH, Ames RW, Wang TK et al. Vitamin D insufficiency and health outcomes over 5 y in older women. The American Journal of Clinical Nutrition 2010 January 1;91(1):82-9.

(158) Putz-Bankuti C, Pilz S, Stojakovic T, Scharnagl H, Pieber TR, Trauner M et al. Association of 25-hydroxyvitamin D levels with liver dysfunction and mortality in chronic liver disease. Liver Int 2012 May 1;32(5):845-51.

(159) Petta S, Cammá C, Scazzone C, Tripodo C, Di Marco V, Bono A et al. Low vitamin D serum level is related to severe fibrosis and low responsiveness to interferon-based therapy in genotype 1 chronic hepatitis C. Hepatology 2010 April 1;51(4):1158-67.

(160) Malham M, Jorgensen SP, Ott P, Agnholt J, Vilstrup H, Borre M et al. Vitamin D deficiency in cirrhosis relates to liver dysfunction rather than aetiology. World J Gastroenterol 2011;17(7):922-5.

(161) Pappa HM, Bern E, Kamin D, Grand RJ. Vitamin D status in gastrointestinal and liver disease. Curr Opin Gastroenterol 2008;24(2):176-83.

(162) Kotsopoulos J, Tworoger SS, Campos H, Chung FL, Clevenger CV, Franke AA et al. Reproducibility of Plasma, Red Blood Cell, and Urine Biomarkers among Premenopausal and Postmenopausal Women from the Nurses' Health Studies. Cancer Epidemiology Biomarkers & Prevention 2010 April 1;19(4):938-46.

(163) Hofmann JN, Yu K, Horst RL, Hayes RB, Purdue MP. Long-term Variation in Serum 25-Hydroxyvitamin D Concentration among Participants in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. Cancer Epidemiology Biomarkers & Prevention 2010 April 1;19(4):927-31.

(164) Bertrand KA, Giovannucci E, Liu Y, Malspeis S, Eliassen AH, Wu K et al. Determinants of plasma 25-hydroxyvitamin D and development of prediction models in three US cohorts. British Journal of Nutrition 2012;108(10):1889-96.

(165) Liu E, Meigs JB, Pittas AG, Economos CD, McKeown NM, Booth SL et al. Predicted 25-hydroxyvitamin D score and incident type 2 diabetes in the Framingham Offspring Study. The American Journal of Clinical Nutrition 2010 June 1;91(6):1627-33.

(166) Wilson PWF, DΓÇÖAgostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. Circulation 1998 May 1;97(18):1837-47.

(167) Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. BMJ 2011 November 28;343.

(168) Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutrition 2002;5(6b):1113-24.

(169) Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. International Journal of Epidemiology 1997 January 1;26(suppl 1):S6.

(170) Slimani N, Deharveng G, Unwin I, Southgate DAT, Vignat J, Skeie G et al. The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. Eur J Clin Nutr 2007 March 21;61(9):1037-56.

(171) Schoenfeld D. Partial residuals for the proportional hazards regression model. Biometrika 1982 April 1;69(1):239-41.

(172) Greenland S, Longnecker MP. Methods for Trend Estimation from Summarized Dose-Response Data, with Applications to Meta-Analysis. American Journal of Epidemiology 1992 June 1;135(11):1301-9.

(173) Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. The American Journal of Clinical Nutrition 1997 April 1;65(4):1220S-8S.

(174) Cole TJ, Kryakin YV. Sympercents: symmetric percentage differences on the 100 logescale simplify the presentation of log transformed data by T. J. Cole, Statistics in Medicine 2000; 19: 3109ΓÇô3125. Statist Med 2002 August 15;21(15):2287-90.

(175) International Agency for Research on Cancer - European Cancer Observatory. 2013.

(176) Helzlsouer KJ, for the VDPP Steering Committee. Overview of the Cohort Consortium Vitamin D Pooling Project of Rarer Cancers. American Journal of Epidemiology 2010 July 1;172(1):4-9.

(177) Skeie G, Braaten T, Hjartaker A, Lentjes M, Amiano P, Jakszyn P et al. Use of dietary supplements in the European Prospective Investigation into Cancer and Nutrition calibration study. Eur J Clin Nutr 0 AD;63(S4):S226-S238.

(178) Willett WC. Nutritional Epidemiology. 2nd edition ed. New York, U.S.: Oxford University Press; 1998.

(179) Chan J, Jaceldo-Siegl K, Fraser G. Determinants of serum 25 hydroxyvitamin D levels in a nationwide cohort of blacks and non-Hispanic whites. Cancer Causes Control 2010;21(4):501-11.

(180) Gilbert R, Martin R, Fraser W, Lewis S, Donovan J, Hamdy F et al. Predictors of 25-hydroxyvitamin D and its association with risk factors for prostate cancer: evidence from the Prostate testing for cancer and Treatment study. Cancer Causes Control 2012;23(4):575-88.

(181) Lips PF, Chapuy MC FAU - Dawson-Hughes, Dawson-Hughes BF, Pols HA FAU, Holick MF. An international comparison of serum 25-hydroxyvitamin D measurements. Osteoporos Int 1999;9(5):394-7.

(182) Wacholder S, Armstrong B, Hartge P. Validation Studies using an Alloyed Gold Standard. American Journal of Epidemiology 1993 June 1;137(11):1251-8.

(183) Spiegelman D, Schneeweiss S, McDermott A. Measurement Error Correction for Logistic Regression Models with an Alloyed Gold Standard. American Journal of Epidemiology 1997 January 15;145(2):184-96.

(184) Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. The American Journal of Clinical Nutrition 2007 September 1;86(3):556-65.

(185) Moghaddam AA, Woodward M, Huxley R. Obesity and Risk of Colorectal Cancer: A Meta-analysis of 31 Studies with 70,000 Events. Cancer Epidemiology Biomarkers & Prevention 2007 December 1;16(12):2533-47.

(186) Wolin KY, Yan Y, Colditz GA, Lee IM. Physical activity and colon cancer prevention: a meta-analysis. Br J Cancer 2009 February 10;100(4):611-6.

(187) Robsahm TE, Aagnes B, Hjartaker A, Langseth H et al. Body mass index, physical activity, and colorectal cancer by anatomical subsites: a systematic review and meta-analysis of cohort studies. Eur J Cancer Prev 2013;(Epub ahead of print).

(188) Hunter D, De Lange M, Snieder H, MacGregor AJ, Swaminathan R, Thakker RV et al. Genetic Contribution to Bone Metabolism, Calcium Excretion, and Vitamin D and Parathyroid Hormone Regulation. J Bone Miner Res 2001 February 1;16(2):371-8.

(189) Engelman CD, Fingerlin TE, Langefeld CD, Hicks PJ, Rich SS, Wagenknecht LE et al. Genetic and Environmental Determinants of 25-Hydroxyvitamin D and 1,25-Dihydroxyvitamin D Levels in Hispanic and African Americans. Journal of Clinical Endocrinology & Metabolism 2008 September 1;93(9):3381-8.

(190) Morton LM, Wang SS, Cozen W, Linet MS, Chatterjee N, Davis S et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. Blood 2008 December 15;112(13):5150-60.

(191) Crosbie EJ, Zwahlen M, Kitchener HC, Egger M, Renehan AG. Body Mass Index, Hormone Replacement Therapy, and Endometrial Cancer Risk: A Meta-Analysis. Cancer Epidemiology Biomarkers & Prevention 2010 December 1;19(12):3119-30.

(192) Janssens W, Bouillon R, Claes B, Carremans C, Lehouck A, Buysschaert I et al. Vitamin D deficiency is highly prevalent in COPD and correlates with variants in the vitamin D-binding gene. Thorax 2010 March 1;65(3):215-20.

(193) Bingham S, Riboli E. Diet and cancer - the European Prospective Investigation into Cancer and Nutrition. Nat Rev Cancer 2004 March;4(3):206-15.

(194) Haftenberger M, Lahmann PH, Panico S, Gonzalez CA, Seidell JC, Boeing H et al. Overweight, obesity and fat distribution in 50- to 64-year-old participants in the European Prospective Investigation into Cancer and Nutrition (EPIC). Public Health Nutrition 2002;5(6b):1147-62.

(195) Blanchard CM, Denniston MM, Baker F, Ainsworth SR, Courneya KS, Hann DM et al. Do Adults Change Their Lifestyle Behaviors After a Cancer Diagnosis? American Journal of Health Behavior 27[3], 246-256. 2003.

(196) Williams K, Steptoe A, Wardle J. Is a cancer diagnosis a trigger for health behaviour change[quest] Findings from a prospective, population-based study. Br J Cancer 2013 June 11;108(11):2407-12.

(197) VanEenwyk J, Campo JS, Ossiander EM. Socioeconomic and demographic disparities in treatment for carcinomas of the colon and rectum. Cancer 2002 July 1;95(1):39-46.

(198) Frederiksen BL, Jorgensen T, Brasso K, Holten I, Osler M. Socioeconomic position and participation in colorectal cancer screening. Br J Cancer 2010 November 9;103(10):1496-501.

(199) Friedenreich C, Norat T, Steindorf K, Boutron-Ruault MC, Pischon T, Mazuir M et al. Physical Activity and Risk of Colon and Rectal Cancers: The European Prospective Investigation into Cancer and Nutrition. Cancer Epidemiology Biomarkers & Prevention 2006 December 1;15(12):2398-407.

(200) Murphy N, Norat T, Ferrari P, Jenab M, Bueno-de-Mesquita B, Skeie G et al. Dietary Fibre Intake and Risks of Cancers of the Colon and Rectum in the European Prospective Investigation into Cancer and Nutrition (EPIC). PLoS ONE 2012 June 22;7(6):e39361.

(201) Jung S, Spiegelman D, Baglietto L, Bernstein L, Boggs DA, van den Brandt PA et al. Fruit and Vegetable Intake and Risk of Breast Cancer by Hormone Receptor Status. Journal of the National Cancer Institute 2013 February 6;105(3):219-36.

(202) Manson JE, Bassuk SS, Lee IM, Cook NR, Albert MA, Gordon D et al. The VITamin D and OmegA-3 TriaL (VITAL): Rationale and design of a large randomized controlled trial of vitamin D and marine omega-3 fatty acid supplements for the primary prevention of cancer and cardiovascular disease. Contemporary Clinical Trials 2012 January;33(1):159-71.

(203) Theodoratou E, Palmer T, Zgaga L, Farrington SM, McKeigue P, Din FVN et al. Instrumental Variable Estimation of the Causal Effect of Plasma 25-Hydroxy-Vitamin D on Colorectal Cancer Risk: A Mendelian Randomization Analysis. PLoS ONE 2012 June 6;7(6):e37662.

(204) Hiraki LT, Qu C, Hutter CM, Baron JA, Berndt SI, Bezieau S et al. Genetic Predictors of Circulating 25-Hydroxyvitamin D and Risk of Colorectal Cancer. Cancer Epidemiology Biomarkers & Prevention 2013 November 1;22(11):2037-46.