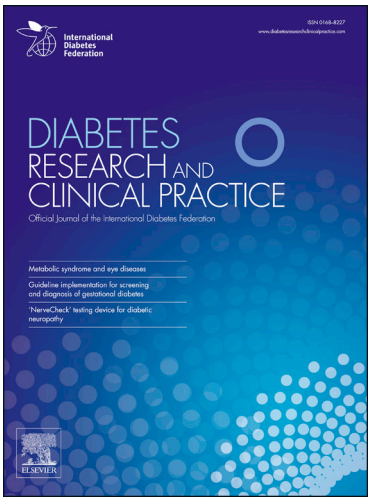


# Vitamin D status and the risk of type 2 diabetes: the Melbourne Collaborative Cohort Study

Alicia K Heath, Elizabeth J Williamson, Allison M Hodge, Peter R Ebeling, Darryl W Eyles, David Kvaskoff, Kerin O'Dea, Graham G Giles, Dallas R English

PII: S0168-8227(17)31927-7
DOI: https://doi.org/10.1016/j.diabres.2018.05.007
Reference: DIAB 7368

To appear in: *Diabetes Research and Clinical Practice*

Received Date: 1 December 2017
Revised Date: 13 April 2018
Accepted Date: 8 May 2018

Please cite this article as: A.K. Heath, E.J. Williamson, A.M. Hodge, P.R. Ebeling, D.W. Eyles, D. Kvaskoff, K. O'Dea, G.G. Giles, D.R. English, Vitamin D status and the risk of type 2 diabetes: the Melbourne Collaborative Cohort Study, *Diabetes Research and Clinical Practice* (2018), doi: https://doi.org/10.1016/j.diabres.2018.05.007

# Vitamin D status and the risk of type 2 diabetes: the Melbourne Collaborative Cohort Study

Alicia K Heath[1,2,3], Elizabeth J Williamson[4,5], Allison M Hodge[1,2], Peter R Ebeling[6], Darryl W Eyles[7,8], David Kvaskoff[7], Kerin O'Dea[1], Graham G Giles[1,2], Dallas R English[1,2]

[1]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, 207 Bouverie St, Melbourne, Victoria 3010, Australia; [2]Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, 615 St Kilda Rd, Melbourne, Victoria 3004, Australia; [3]Nuffield Department of Population Health, University of Oxford, Roosevelt Drive, Oxford OX3 7LF, UK; [4]Farr Institute of Health Informatics Research, 222 Euston Rd, Kings Cross, London NW1 2DA, UK; [5]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, Keppel St, Bloomsbury, London WC1E 7HT, UK; [6]Department of Medicine, School of Clinical Sciences, Monash University, Monash Medical Centre, Clayton, Victoria 3168, Australia; [7]Queensland Brain Institute, University of Queensland, St Lucia, Queensland 4072, Australia; [8]Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Wacol, Queensland 4076, Australia

Corresponding author: Professor Dallas English, Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, Level 3, 207 Bouverie Street, The University of Melbourne, Victoria 3010, Australia. Telephone: +61 3 8344 0742, email: d.english@unimelb.edu.au

**Word count (including main text, tables and figure legends): 4783 words**

**3 tables and 2 figures**

**1 supplementary table**

**Abstract**

**Aims**

Inverse associations between vitamin D status and risk of type 2 diabetes observed in epidemiological studies could be biased by confounding and reverse causality. We investigated the prospective association between vitamin D status and type 2 diabetes and the possible role of reverse causality.

**Methods**

We conducted a case-cohort study within the Melbourne Collaborative Cohort Study (MCCS), including a random sample of 628 participants who developed diabetes and a sex-stratified random sample of the cohort (*n*=1,884). Concentration of 25-hydroxyvitamin D (25(OH)D) was measured using liquid chromatography-tandem mass spectrometry in samples collected at recruitment. Logistic regression was used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the risk of type 2 diabetes for quartiles of 25(OH)D relative to the lowest quartile and per 25 nmol/L increase in 25(OH)D, adjusting for confounding variables.

**Results**

The ORs for the highest versus lowest 25(OH)D quartile and per 25 nmol/L increase in 25(OH)D were 0.60 (95% CI: 0.44, 0.81) and 0.76 (95% CI: 0.63, 0.92; *p*=0.004), respectively. In participants who reported being in good/very good/excellent health approximately four years after recruitment, ORs for the highest versus lowest 25(OH)D quartile and per 25 nmol/L increase in 25(OH)D were 0.46 (95% CI: 0.29, 0.72) and 0.71 (95% CI: 0.56, 0.89; *p*=0.003), respectively.

**Conclusions**

In this sample of middle-aged Australians, vitamin D status was inversely associated with the risk of type 2 diabetes, and this association did not appear to be explained by reverse causality.

**Keywords**: Vitamin D; 25-hydroxyvitamin D; Vitamin D deficiency; Type 2 diabetes mellitus

## 1 Introduction

Globally the prevalence of type 2 diabetes is increasing and if current trends continue, more than 642 million people (10% of adults) will have diabetes by 2040 [1]. Well-established risk factors for type 2 diabetes include excess body weight (particularly abdominal adiposity), physical inactivity, poor diet, increasing age, family history of type 2 diabetes, ethnicity, and genetics [1]. Identifying other risk factors could inform strategies for prevention.

Vitamin D has been linked to the pathogenesis of type 2 diabetes [2-7]. Several meta-analyses of prospective studies have found an inverse association between vitamin D status, as assessed by circulating serum or plasma 25-hydroxyvitamin D (25(OH)D) concentrations, and risk of type 2 diabetes [8, 9]. However, these results have not been replicated in randomised controlled trials (RCTs) of vitamin D supplementation [10], and a recent umbrella review of the literature pertaining to vitamin D and multiple health outcomes concluded that there was only suggestive evidence of an association between vitamin D and type 2 diabetes [11]. Existing RCTs have been criticised for design limitations such as lack of statistical power and inclusion of vitamin D replete individuals [12, 13], while results from observational studies could be biased by confounding and reverse causality. Confounding is possible because vitamin D status is associated with several risk factors for diabetes (such as obesity, physical inactivity, age and ethnicity). Reverse causality would occur if study participants were in poor health at study entry due to undiagnosed diabetes, and this led to poor vitamin D status (for example via reduced sun exposure, dietary changes, or increased inflammation). Existing prospective cohort studies have not extensively explored the possibility of reverse causality, engendering uncertainty regarding any potential benefit of vitamin D for the prevention of type 2 diabetes.

We investigated the association between vitamin D status and the risk of type 2 diabetes in a large population-based prospective cohort study and examined whether this association is likely to be explained by reverse causality.

## 2 Subjects, Materials and Methods

A case-cohort study to investigate vitamin D status and the risk of cancer, type 2 diabetes and mortality was conducted within the Melbourne Collaborative Cohort Study (MCCS). The MCCS is a prospective cohort study of 41,514 participants (24,469 women and 17,045 men) living in the Melbourne metropolitan area who were predominantly aged between 40 and 69 years at study recruitment (1990-1994). Southern European migrants were deliberately recruited (approximately 25% of the cohort) to extend the range of dietary and lifestyle exposures. Details of the MCCS have been published [14]. Briefly, at baseline (wave 1) extensive demographic, lifestyle and dietary data were collected, and anthropometric measurements were performed. Participants were also asked whether they had ever been diagnosed with diabetes mellitus. Blood samples were collected, from which plasma glucose concentrations (67% fasting) were measured using Kodak Ektachem DT60 desktop analysers (Rochester, NY). From one year into study recruitment (for approximately 75% of participants), whole blood was spotted onto Guthrie cards, which were air dried and stored in dark conditions.

Approximately four years after baseline (wave 2), participants were mailed a self-administered questionnaire which asked about non-fatal and non-cancer health events, including diabetes. For self-reported incident cases of diabetes, 76% were confirmed by their GP as having type 2 diabetes [15]. Participants were also asked: “In general, would you say your health is: excellent, very good, good, fair, or poor?”. A third wave of data collection was

conducted between 2003 and 2007, when participants attended a clinic where further questionnaires were completed, anthropometric measurements, medication use (including insulin and oral hypoglycemics), and several disease endpoints were recorded, and another blood sample was collected. Plasma glucose concentrations were measured using a glucometer. Self-reported diabetes status and year of diagnosis were also recorded.

The Cancer Council Victoria's Human Research Ethics Committee approved the study protocol and participants gave written consent to participate.

### 2.1 Participants

Eligibility for the case-cohort study was restricted to the 29,206 participants who had no cancer diagnosis before baseline and for whom dried blood spots were available from baseline blood samples. For the diabetes component, we excluded 132 people with pre-existing diabetes or unknown diabetes status at baseline, where diabetes status was determined from self-report or plasma glucose concentrations (see Table 1), leaving 29,074 eligible.

The diabetes study included a random sample of participants who developed diabetes between baseline and wave 3, and a sex-stratified random sample of all eligible participants ('subcohort'). Participants for whom 25(OH)D measurements were not performed, with missing data for any confounding variable, or with extreme total energy intakes ($<1^{st}$ and $>99^{th}$ sex-specific percentiles) were excluded. Participants with missing data on diabetes status were excluded from analyses because their case status was unknown.

### 2.2 Assessment of vitamin D status

Concentrations of 25(OH)$D_2$ and 25(OH)$D_3$ from baseline dried blood spot samples were measured using liquid chromatography-tandem mass spectrometry and summed to give total 25(OH)D [16]. Samples were processed in random order and laboratory analysts were blind to outcome status of participants. Results are presented for total 25(OH)D concentrations. Analyses were also conducted using 25(OH)$D_3$ but there was no material difference in results, as few participants had any circulating 25(OH)$D_2$, which also precluded separate analysis of 25(OH)$D_2$. To remove batch effects, a mixed-effects linear regression model with a random effect for batch was fitted for 25(OH)D levels of subcohort participants, then for all participants, the predicted batch-specific deviations from the overall mean were subtracted from the observed values. Concentrations of 25(OH)D exhibited sinusoidal seasonal variation, which was removed using trigonometric regression [17]. Concentrations of 25(OH)D are reported as plasma equivalents, obtained using a previously developed calibration equation [18]. Participants were divided into sex-specific quartiles based on the distribution of batch- and season-adjusted plasma 25(OH)D for the subcohort.

### 2.3 Ascertainment of diabetes cases

Diabetes status was assessed at baseline, wave 2 and wave 3 using the criteria shown in Table 1. On each occasion, participants who satisfied any of the criteria were classified as having diabetes, while those with incomplete information were classified as having missing diabetes status. The World Health Organization diagnostic criteria for plasma glucose concentrations indicative of diabetes were used [19]. No distinction between type 1 and type 2 diabetes was made, however, all incident cases were assumed to be type 2 because this is most likely after the age of 40 years [15, 20]. Classification of diabetes status at wave 3 used the same criteria as at baseline, except that where diabetes status was missing, participants using any diabetes

medication were classified as a case, while those who did not report using any diabetes medications were considered not to have diabetes. A participant was classified as an incident case if they did not have diabetes at baseline and were identified as a case at either wave 2 or wave 3.

### 2.4 Confounders

The following confounding variables measured at baseline were included in analyses based on *a priori* knowledge and use of a causal diagram: sex, age (six categories: <45, 45–49, 50–54, 55–59, 60–64, and ≥65 years), country of birth (Australia/New Zealand/Northern Europe or Southern Europe), an area-based measure of socio-economic status (Socio-Economic Indexes for Areas (SEIFA); quintiles from most disadvantaged to least disadvantaged), highest education level attained (primary school, some secondary school, secondary school, and tertiary qualification), alcohol consumption (five categories: never, former, and sex-specific tertiles of current intake), smoking status (never, former, current), physical activity (four categories reflecting the frequency and intensity of recreational activity in the past 6 months), waist circumference (sex-specific quartiles, cm), Mediterranean diet score (three categories, with the highest indicating high adherence to a Mediterranean dietary pattern), margarine intake (quartiles, times/week), total energy intake (sex-specific quartiles, kJ/day), history of hypertension at baseline, and history of cardiovascular disease (CVD; includes history of angina, myocardial infarction, or stroke) at baseline. Margarine intake was included because margarine is fortified with vitamin D in Australia and it was associated with circulating 25(OH)D concentration in this cohort (data not shown), and because polyunsaturated or monounsaturated fats (which are present in margarine) have been reported to be associated with the risk of type 2 diabetes [21, 22].

### 2.5 Statistical analyses

Statistical analyses were performed using Stata version 14.1 (StataCorp, College Station, Texas, USA).

In light of the unconventional design of this study, with ascertainment of case status at two time points, there was no established strategy to use for analysing the data. Two strategies were explored, with no material difference in their results.

All results presented are from an analysis strategy that resembled a nested case-control study with density sampling. Three controls per case were selected (without replacement) at the same time as cases were identified and matched on sex. The flow diagram of participants included in the main analyses and their case status at each wave is shown in Figure 1. There were 83 women and 77 men (total 160) with diabetes at wave 2. A random sample of 249 female and 231 male controls (total 480) was selected from the 2,391 subcohort participants who completed the wave 2 questionnaire and did not have diabetes at wave 2 (regardless of whether they later developed diabetes). There were 231 women and 237 men (total 468) with incident diabetes identified at wave 3 (i.e. who were not cases at wave 2). A sample of 693 female and 711 male controls (total 1,404) was randomly selected from the 1,491 subcohort participants who were not selected as controls at wave 2, attended the wave 3 clinic, and had not developed diabetes before wave 3.

The alternative analysis strategy was based on a nested case-control study with selection of controls at the end of follow-up (sometimes referred to as cumulative sampling). Cases were all people with diabetes (randomly selected cases plus subcohort cases) regardless of when

they were identified as having diabetes (at wave 2 or wave 3). Controls were all members of the subcohort who attended the wave 3 clinic, and had not developed diabetes before wave 3 (n=1,827). Results from this approach are presented in the Supplementary material.

Logistic regression was used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the risk of diabetes for each quartile of 25(OH)D relative to the lowest quartile, with adjustment for the confounders listed above. Test for trend across categories was performed by including categorical 25(OH)D as a linear term in the model. The OR per 25 nmol/L increase in 25(OH)D was estimated by using the continuous form of this variable. Potential non-linearity of the dose-response relationship was explored by fitting a restricted cubic spline model with four knots at fixed and equally-spaced percentiles (5%, 35%, 65%, 95%) of 25(OH)D.

#### 2.5.1 Interaction by time since baseline

An interaction was fitted between 25(OH)D and time since baseline, where time=0 for wave 2 cases and controls, and time=1 for wave 3 cases and controls. This analysis was designed to assess whether the association differed by time since baseline (using two time points as it was not possible to assess time continuously).

#### 2.5.2 Effect modification by sex and baseline disease status

To explore possible effect modification by sex, an interaction was fitted between continuous 25(OH)D and sex. Effect modification by baseline disease status was assessed by fitting interactions between continuous 25(OH)D and history of hypertension and history of CVD.

### 2.5.3 Sensitivity analysis

To investigate potential reverse causality, a sensitivity analysis was performed restricted to the 2,281 participants who reported being in good, very good, or excellent health at wave 2 and who did not have diabetes at wave 2. Of these, 139 women and 159 men (total 298) had diabetes at wave 3. A random sample of 417 female and 477 male controls (total 894) was selected from subcohort participants who did not have diabetes at wave 3.

## 3 Results

### 3.1 Participants

Of the 3,408 participants selected for the diabetes case-cohort study, 25(OH)D measurements were not performed for 13 participants, confounder data was missing for 17 participants, and 55 had extreme values for daily total energy intake. After exclusion of these participants, 3,323 were eligible for analysis. In total, 628 people with diabetes were included in the analyses, of whom 109 were in the subcohort and 519 were non-subcohort cases. Controls for the nested density-sampled case-control study (for which results are presented) comprised 480 subcohort participants who did not develop diabetes before wave 2 (eight of whom later developed diabetes), and a further 1,404 subcohort participants who were selected as controls at wave 3, as shown in Figure 1. Characteristics of diabetes cases and these subcohort controls are shown in Table 2.

### 3.2 Main results

The OR for the highest compared with the lowest 25(OH)D quartile was 0.60 (95% CI: 0.44, 0.81) and the OR per 25 nmol/L increase in 25(OH)D was 0.76 (95% CI: 0.63, 0.92; $p$=0.004). Results were similar for 25(OH)$D_3$ (data not shown). There was evidence that the spline model fitted better than a linear trend ($p$=0.003), with a sharp reduction in risk as

25(OH)D increased from 40 to 60 nmol/L, followed by a plateau at higher 25(OH)D concentrations (Figure 2). Results from the alternative analysis strategy were almost identical (Supplementary Table 1).

#### 3.2.1 Interaction by time since baseline

The average time between baseline (wave 1) and wave 2 was 4.0 (SD, 0.4) years (maximum 6.8 years) and the average time between baseline and wave 3 was 11.4 (SD, 1.2) years (maximum follow-up time=14.6 years). The ORs were similar for the two strata defined by follow-up period (Table 3). For the quartile analysis, *p* for time interaction=0.69 and for the continuous analysis, *p* for time interaction=0.42.

#### 3.2.2 Effect modification by sex and baseline disease status

There was no evidence that the association varied by sex (*p* for interaction=0.49). There was no evidence of interaction between vitamin D status and history of hypertension at baseline (*p* for interaction=0.35), based on 505 participants (213 cases) with a history pf hypertension. Similarly, there was no evidence of interaction between vitamin D status and history of CVD at baseline (*p* for interaction=0.79), based on 145 participants (54 cases) with a history of CVD.

### 3.3 Sensitivity analysis: association between vitamin D status and diabetes for people in good to excellent health at wave 2

The sensitivity analysis included 298 diabetes cases and 894 controls (total 1,192 participants). The association for these participants was slightly stronger than for all participants. The OR for the highest compared with the lowest 25(OH)D quartile was 0.46;

95% CI: 0.29, 0.72; $p_{trend}$<0.001) and the OR per 25 nmol/L increase in 25(OH)D was 0.71 (95% CI: 0.56, 0.89; $p$=0.003).

## 4 Discussion

Vitamin D status was inversely associated with the incidence of type 2 diabetes over 11 years of follow-up. Each 25 nmol/L increment in 25(OH)D was associated with an approximately 24% lower risk of type 2 diabetes after controlling for well-known risk factors, with some evidence that the dose-response curve plateaued at higher concentrations. The association was slightly stronger for participants who reported being in good to excellent health approximately 4 years after baseline; these participants had a 29% lower risk of type 2 diabetes per 25 nmol/L increment in 25(OH)D during an interval approximately 4 to 11 years after baseline.

Strengths of this study include the prospective design, long follow-up, large number of cases, extensive data on potential confounders, and generalisability based on a broad age range and community-based recruitment. A major strength was the availability of data on general health status several years after blood samples were collected, which facilitated assessment of reverse causality. Limitations included the different methods used to assess diabetes status at each wave of follow-up, and that participants were not specifically asked whether they had type 1 or type 2 diabetes. Nevertheless, it is unlikely that participants developed type 1 diabetes during follow-up because this is usually diagnosed before the age of 40 years [20]. Few participants in this study had circulating 25(OH)$D_2$ which precluded assessment of the association between 25(OH)$D_2$ and diabetes. Reported 25(OH)D concentrations should be interpreted cautiously as these were plasma-equivalent concentrations estimated from dried blood spots and adjusted for batch and season [18]. A potential source of bias was selection

bias due to people not participating in wave 2 or wave 3. Selection bias would occur if participation was a common "effect" of 25(OH)D concentrations and the outcome. However, for the subcohort, inclusion in the analyses was not strongly associated with 25(OH)D concentrations (data not shown), suggesting that selection bias is unlikely to explain the observed association between 25(OH)D and risk of diabetes. The MCCS only included participants of European descent, thus the results cannot necessarily be extrapolated to other ethnicities. Because diabetes status was assessed at two different time points, and there are no definitive analysis methods for this unconventional case-cohort study design, two separate analysis strategies were employed, and there was no material difference in results between them. It is therefore unlikely that the results were biased by the analysis strategy. While adjustments were made for known confounders, the possibility of residual confounding cannot be ruled out. Nevertheless, analyses from the National Health and Nutrition Examination Surveys (NHANES) showing consistent inverse associations between 25(OH)D concentrations and diabetes risk among non-Hispanic whites and Mexican-Americans, but not among non-Hispanic blacks [23], suggest that the association between 25(OH)D and diabetes is unlikely to be due to inadequate control for confounding, which might be expected to work the same way across the ethnic groups.

Despite consistent evidence from observational studies of an inverse association between 25(OH)D concentrations and incident type 2 diabetes [8, 9], there is no evidence from RCTs to support a causal association [10]. A possible explanation for the null results from RCTs is that vitamin D sufficiency might need to be sustained over long periods to have any benefit. It is also possible that vitamin D sufficiency might need to be maintained throughout the entire lifetime, and supplementation may not be able to reverse disease processes once they are initiated [13, 24]. The results from this study, in which the association did not markedly

change over time, and was stronger for participants in good to excellent health some years after blood sampling, support the notion that adequate vitamin D is required to reduce the risk of diseases that progress over a period of several years, and that vitamin D adequacy may be required long before the disease process is established.

A Mendelian randomisation study of common genetic variants related to 25(OH)D synthesis and metabolism found that 7-dehydrocholesterol reductase (DHCR7) variants associated with low plasma 25(OH)D concentration were associated with an increased risk of diabetes (OR for a genetically determined 20 nmol/L lower 25(OH)D=1.51; 95% CI: 0.98, 2.33; $p_{trend}$=0.04 for type 2 diabetes and 1.54; 95% CI: 1.03, 2.30 for any diabetes) [25]. The same study showed no significant associations between CYP2R1 variants or allele scores and the risk of diabetes (OR for a genetically determined 20 nmol/L lower 25(OH)D=1.02; 95% CI: 0.75, 1.37; $p_{trend}$=0.84 for type 2 diabetes and 1.01; 95% CI: 0.76, 1.35 for any diabetes) [25]. Another Mendelian randomisation study did not find a statistically significant relationship between genetically low 25(OH)D (using four genetic variants) and the risk of type 2 diabetes (OR per 25 nmol/L lower 25(OH)D=0.93; 95% CI: 0.77, 1.13; $p$=0.46); this association was in the opposite direction to that reported in the aforementioned Mendelian randomisation study, and the authors concluded that there was inadequate evidence to support a causal relationship [26]. Taken together, these findings suggest that reverse causality might explain the results from observational studies.

The possibility of reverse causality has been a limitation of existing observational studies investigating the association between vitamin D status and disease. This is of particular concern for an outcome such as diabetes, for which people can remain asymptomatic and undiagnosed for years [1]. It is possible that lifestyle changes (e.g. in diet and outdoor

activity) or suboptimal health prior to diagnosis, for example increased inflammation [27], or hyperglycaemia causing tissue damage in undiagnosed diabetes, could provoke a reduction in 25(OH)D concentrations. If the association was due to reverse causality then a much stronger association would be expected to be observed in the first few years of follow-up. In the MCCS, the association at wave 2 (approximately 4 years after baseline) was similar to the association at wave 3 (approximately 11 years after baseline). The persistence of the association over time suggests that reverse causality is an unlikely explanation for the association observed in this study. However, due to the limited number of cases at wave 2 (*n*=160), any potential interaction with time could not be explored in depth. The sensitivity analysis restricted to participants who reported being in good to excellent health permitted a more thorough exploration of whether reverse causality could explain the observed association. The association was slightly stronger for participants who were in good/very good/excellent health approximately 4 years after baseline. For these participants, the risk of diabetes was approximately 54% lower for those with the highest compared with the lowest 25(OH)D concentrations. These findings imply that the association between vitamin D status and diabetes in this study is unlikely to be due to reverse causality.

The Australian Diabetes, Obesity and Lifestyle (AusDiab) study reported that each 25 nmol/L increment in 25(OH)D was associated with a 24% lower risk of developing type 2 diabetes (OR = 0.76, 95% CI: 0.63, 0.92) [28]. The results of this small study (199 diabetes cases diagnosed during 5 years of follow-up) are consistent with our findings from the MCCS, providing strong evidence that vitamin D insufficiency is associated with an increased risk of type 2 diabetes among Australians.

Several reviews have outlined potential mechanisms for a role of vitamin D in the pathogenesis of type 2 diabetes [2-7]. Vitamin D might contribute to type 2 diabetes by influencing insulin secretion and sensitivity [4, 5, 29]. The function of pancreatic beta cells, which produce insulin, appears to be influenced by vitamin D. In particular, the active form of vitamin $D_3$, 1,25-dihydroxyvitamin $D_3$ [$1,25(OH)_2D_3$] is believed to be important for insulin synthesis and secretion [2-4, 29-31]. Pancreatic beta cells possess vitamin D receptors (VDRs) [32], and express the 1α-hydroxylase enzyme which converts $25(OH)D_3$ to the active form [$1,25(OH)_2D_3$] [33]. Vitamin D is also thought to influence insulin secretion via extracellular and intracellular calcium levels, which govern release of insulin from beta cells [2, 3, 29, 34]. Tissues involved in the development of type 2 diabetes, such as adipose tissue and skeletal muscle, have VDRs, and locally-produced $1,25(OH)_2D_3$ in these tissues increases insulin sensitivity [7]. Vitamin D could also contribute to insulin sensitivity by regulating extracellular calcium, calcium influx, and intracellular calcium concentrations required for insulin-mediated functions such as glucose transport [2]. In addition, a vitamin D response element (VDRE) is present in the promoter region of the insulin receptor gene [35], and $1,25(OH)_2D_3$ activates expression of this gene [36]. Consistent with experimental studies, observational studies have found an inverse association between 25(OH)D concentrations and insulin resistance [28, 37-39]. Finally, vitamin D could also indirectly contribute to the pathogenesis of type 2 diabetes via regulation of inflammatory processes (such as production of cytokines) associated with insulin resistance and beta cell death [2, 7, 40].

Overall, a putative role of vitamin D in the aetiology of type 2 diabetes appears to be biologically plausible. In the MCCS, vitamin D status was inversely associated with the incidence of type 2 diabetes, and this association did not appear to be explained by reverse causality. Further long-term intervention studies in vitamin D deficient people at risk for

diabetes are required to confirm whether vitamin D is causally associated with type 2 diabetes.

**Acknowledgements**

This study was made possible by the contribution of many people, including the original investigators, the Program Manager, and the diligent team who recruited the participants and who continue working on follow-up. We would like to express our gratitude to the many thousands of Melbourne residents who continue to participate in the study. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database.

**Funding**

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian National Health and Medical Research Council (NHMRC) grants, 209057 and 396414 and by infrastructure provided by Cancer Council Victoria. This project was specifically funded by the National Health and Medical Research Council project grant 623208. Alicia Heath is supported by an Australian Postgraduate Award. The funding sources had no involvement in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Declarations of interest: none.

**Contribution statement**

AKH conducted the literature search, designed the study, collected, analysed and interpreted the data, and wrote the manuscript. EJW oversaw the study design, statistical analysis, and interpretation of the data. DK performed laboratory measurements of vitamin D status. AMH

and PRE were involved in the design of the study and review of the manuscript. GGG conceived and designed the MCCS and reviewed the final manuscript. DWE obtained funding, oversaw measurements of vitamin D status and reviewed the final report. DRE obtained funding, oversaw the design of the study, data analysis, interpretation of results, and drafting of the report. All authors read and approved the final manuscript. DRE had full access to all the data in the study and had final responsibility for the decision to submit for publication.

**References**

[1] International Diabetes Federation. IDF Diabetes Atlas, Seventh Edition. Brussels, Belgium; 2015.

[2] Pittas AG, Lau J, Hu FB, Dawson-Hughes B. Review: The role of vitamin D and calcium in type 2 diabetes. A systematic review and meta-analysis. J Clin Endocrinol Metab. 2007;92:2017-29.

[3] Palomer X, Gonzalez-Clemente JM, Blanco-Vaca F, Mauricio D. Role of vitamin D in the pathogenesis of type 2 diabetes mellitus. Diabetes Obes Metab. 2008;10:185-97.

[4] Tai K, Need AG, Horowitz M, Chapman IM. Vitamin D, glucose, insulin, and insulin sensitivity. Nutrition. 2008;24:279-85.

[5] Alvarez JA, Ashraf A. Role of vitamin D in insulin secretion and insulin sensitivity for glucose homeostasis. Int J Endocrinol. 2010;2010:351385.

[6] Stivelman E, Retnakaran R. Role of vitamin D in the pathophysiology and treatment of type 2 diabetes. Curr Diabetes Rev. 2012;8:42-7.

[7] Mathieu C. Vitamin D and diabetes: where do we stand? Diabetes Res Clin Pract. 2015;108:201-9.

[8] Song Y, Wang L, Pittas AG, Del Gobbo LC, Zhang C, Manson JE, et al. Blood 25-hydroxy vitamin D levels and incident type 2 diabetes: a meta-analysis of prospective studies. Diabetes Care. 2013;36:1422-8.

[9] Ekmekcioglu C, Haluza D, Kundi M. 25-Hydroxyvitamin D status and risk for colorectal cancer and type 2 diabetes mellitus: A systematic review and meta-analysis of epidemiological studies. Int J Environ Res Public Health. 2017;14:127.

[10] Seida JC, Mitri J, Colmers IN, Majumdar SR, Davidson MB, Edwards AL, et al. Effect of vitamin $D_3$ supplementation on improving glucose homeostasis and preventing diabetes: a systematic review and meta-analysis. J Clin Endocrinol Metab. 2014;99:3551-60.

[11] Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JP. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ. 2014;348:g2035.

[12] Jorde R, Grimnes G. Vitamin D and health: the need for more randomized controlled trials. J Steroid Biochem Mol Biol. 2015;148:269-74.

[13] Lappe JM, Heaney RP. Why randomized controlled trials of calcium and vitamin D sometimes fail. Dermatoendocrinol. 2012;4:95-100.

[14] Milne RL, Fletcher AS, MacInnis RJ, Hodge AM, Hopkins AH, Bassett JK, et al. Cohort profile: the Melbourne Collaborative Cohort Study (Health 2020). Int J Epidemiol. 2017;46:1757-i.

[15] Hodge AM, English DR, O'Dea K, Giles GG. Glycemic index and dietary fiber and the risk of type 2 diabetes. Diabetes Care. 2004;27:2701-6.

[16] Eyles D, Anderson C, Ko P, Jones A, Thomas A, Burne T, et al. A sensitive LC/MS/MS assay of 25OH vitamin $D_3$ and 25OH vitamin $D_2$ in dried blood spots. Clin Chim Acta. 2009;403:145-51.

[17] Cox NJ. Speaking Stata: In praise of trigonometric predictors. The Stata Journal. 2006;6:561-79.

[18] Heath AK, Williamson EJ, Ebeling PR, Kvaskoff D, Eyles DW, English DR. Measurements of 25-hydroxyvitamin D concentrations in archived dried blood spots are reliable and accurately reflect those in plasma. J Clin Endocrinol Metab. 2014;99:3319-24.

[19] World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: report of a WHO/IDF consultation. Geneva, Switzerland; 2006.

[20] Australian Institute of Health and Welfare. How many Australians have diabetes?, http://www.aihw.gov.au/how-common-is-diabetes/ [accessed 13 February 2016].

[21] Risérus U, Willett WC, Hu FB. Dietary fats and prevention of type 2 diabetes. Prog Lipid Res. 2009;48:44-51.

[22] Hodge AM, English DR, O'Dea K, Sinclair AJ, Makrides M, Gibson RA, et al. Plasma phospholipid and dietary fatty acids as predictors of type 2 diabetes: interpreting the role of linoleic acid. Am J Clin Nutr. 2007;86:189-97.

[23] Christensen MHE, Scragg RK. Consistent ethnic specific differences in diabetes risk and vitamin D status in the National Health and Nutrition Examination Surveys. Journal of Steroid Biochemistry and Molecular Biology. 2016;164:4-10.

[24] Gillie O. Controlled trials of vitamin D, causality and type 2 statistical error. Public Health Nutr. 2016;19:409-14.

[25] Afzal S, Brøndum-Jacobsen P, Bojesen SE, Nordestgaard BG. Vitamin D concentration, obesity, and risk of diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2014;2:298-306.

[26] Ye Z, Sharp SJ, Burgess S, Scott RA, Imamura F, InterAct Consortium, et al. Association between circulating 25-hydroxyvitamin D and incident type 2 diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2015;3:35-42.

[27] Autier P, Boniol M, Pizot C, Mullie P. Vitamin D status and ill health: a systematic review. Lancet Diabetes Endocrinol. 2014;2:76-89.

[28] Gagnon C, Lu ZX, Magliano DJ, Dunstan DW, Shaw JE, Zimmet PZ, et al. Serum 25-hydroxyvitamin D, calcium intake, and risk of type 2 diabetes after 5 years: results from a national, population-based prospective study (the Australian Diabetes, Obesity and Lifestyle study). Diabetes Care. 2011;34:1133-8.

[29] Takiishi T, Gysemans C, Bouillon R, Mathieu C. Vitamin D and diabetes. Endocrinol Metab Clin N Am. 2010;39:419-46.

[30] Norman AW, Frankel JB, Heldt AM, Grodsky GM. Vitamin D deficiency inhibits pancreatic secretion of insulin. Science. 1980;209:823-5.

[31] Clark SA, Stumpf WE, Sar M. Effect of 1,25 dihydroxyvitamin $D_3$ on insulin secretion. Diabetes. 1981;30:382-6.

[32] Johnson JA, Grande JP, Roche PC, Kumar R. Immunohistochemical localization of the $1,25(OH)_2D_3$ receptor and calbindin $D_{28k}$ in human and rat pancreas. Am J Physiol. 1994;267:E356-60.

[33] Bland R, Markovic D, Hills CE, Hughes SV, Chan SL, Squires PE, et al. Expression of 25-hydroxyvitamin $D_3$-1a-hydroxylase in pancreatic islets. J Steroid Biochem Mol Biol. 2004;89-90:121-5.

[34] Wollheim CB, Sharp GW. Regulation of insulin release by calcium. Physiol Rev. 1981;61:914-73.

[35] Maestro B, Dávila N, Carranza MC, Calle C. Identification of a Vitamin D response element in the human insulin receptor gene promoter. J Steroid Biochem Mol Biol. 2003;84:223-30.

[36] Maestro B, Molero S, Bajo S, Davila N, Calle C. Transcriptional activation of the human insulin receptor gene by 1,25-dihydroxyvitamin $D_3$. Cell Biochem Funct. 2002;20:227-32.

[37] Forouhi NG, Luan J, Cooper A, Boucher BJ, Wareham NJ. Baseline serum 25-hydroxy vitamin D is predictive of future glycemic status and insulin resistance: the Medical Research Council Ely Prospective Study 1990-2000. Diabetes. 2008;57:2619-25.

[38] Girgis CM, Clifton-Bligh RJ, Turner N, Lau SL, Gunton JE. Effects of vitamin D in skeletal muscle: falls, strength, athletic performance and insulin sensitivity. Clin Endocrinol. 2014;80:169-81.

[39] Kayaniyil S, Vieth R, Retnakaran R, Knight JA, Qi Y, Gerstein HC, et al. Association of vitamin D with insulin resistance and beta-cell dysfunction in subjects at risk for type 2 diabetes. Diabetes Care. 2010;33:1379-81.

[40] Chagas CE, Borges MC, Martini LA, Rogero MM. Focus on vitamin D, inflammation and type 2 diabetes. Nutrients. 2012;4:52-67.

**Table 1: Criteria for diabetes mellitus at baseline and each wave of follow-up.**

| Time point | Definition of diabetes mellitus |
|---|---|
| Baseline/Wave 1 (1990 – 1994) | Fasting plasma glucose ≥ 7.0 mmol/L, or<br>Non-fasting plasma glucose ≥ 11.1 mmol/L, or<br>Self-reported diabetes |
| Wave 2 (1995 – 2002) | Self-reported diabetes |
| Wave 3 (2003 – 2007) | Fasting plasma glucose ≥ 7.0 mmol/L, or<br>Non-fasting plasma glucose ≥ 11.1 mmol/L, or<br>Self-reported diabetes, or<br>Using diabetes medication |

**Table 2: Characteristics of diabetes cases and subcohort participants without diabetes.**

| | Subcohort non-cases | Diabetes cases |
|---|---|---|
| *N* | 1,884 | 628 |
| 25(OH)D (nmol/L), median (IQR) | 49.7 (23.4) | 44.6 (22.6) |
| Sex | | |
| Female | 942 (50.0) | 314 (50.0) |
| Male | 942 (50.0) | 314 (50.0) |
| Age (years), median (IQR) | 53.1 (14.7) | 55.8 (13.3) |
| Country of birth | | |
| Australia/New Zealand/Northern Europe | 1,649 (87.5) | 457 (72.8) |
| Southern Europe | 235 (12.5) | 171 (27.2) |
| Socioeconomic disadvantage | | |
| 1st quintile (most disadvantage) | 210 (11.1) | 108 (17.2) |
| 2nd quintile | 274 (14.5) | 118 (18.8) |
| 3rd quintile | 299 (15.9) | 116 (18.5) |
| 4th quintile | 453 (24.0) | 132 (21.0) |
| 5th quintile (least disadvantage) | 648 (34.4) | 154 (24.5) |
| Educational attainment | | |
| Primary school or less | 157 (8.3) | 124 (19.7) |
| Some secondary school | 687 (36.5) | 267 (42.5) |
| Secondary school | 452 (24.0) | 126 (20.1) |
| Tertiary qualification | 588 (31.2) | 111 (17.7) |
| Alcohol intake (g/day) | | |
| Never | 401 (21.3) | 194 (30.9) |
| Former | 70 (3.7) | 34 (5.4) |
| Current low | 451 (23.9) | 150 (23.9) |
| Current medium | 461 (24.5) | 126 (20.1) |
| Current high | 501 (26.6) | 124 (19.7) |
| Smoking | | |
| Never | 1,085 (57.6) | 351 (55.9) |
| Former | 632 (33.5) | 205 (32.6) |
| Current | 167 (8.9) | 72 (11.5) |
| Physical activity | | |
| None | 356 (18.9) | 173 (27.5) |
| Low | 370 (19.6) | 126 (20.1) |
| Moderate | 618 (32.8) | 219 (34.9) |
| High | 540 (28.7) | 110 (17.5) |
| Waist circumference (cm, quartiles)[a] | | |
| 1 | 526 (27.9) | 24 (3.8) |
| 2 | 487 (25.8) | 68 (10.8) |
| 3 | 499 (26.5) | 157 (25.0) |
| 4 | 372 (19.7) | 379 (60.4) |
| Mediterranean diet score | | |
| 0 – 3 (low) | 428 (22.7) | 165 (26.3) |
| 4 – 6 (moderate) | 1,191 (63.2) | 395 (62.9) |
| 7 – 9 (high) | 265 (14.1) | 68 (10.8) |
| Margarine intake (times/week, quartiles) | | |
| 0.0 – 0.4 | 354 (18.8) | 118 (18.8) |
| 0.5 – 6.9 | 514 (27.3) | 182 (29.0) |
| 7.0 – 17.4 | 445 (23.6) | 140 (22.3) |
| ≥ 17.5 | 571 (30.3) | 188 (29.9) |
| Energy intake (kJ/day, quartiles)[b] | | |
| 1 | 422(22.4) | 176 (28.0) |
| 2 | 503 (26.7) | 155 (24.7) |
| 3 | 490 (26.0) | 132 (21.0) |
| 4 | 469 (24.9) | 165 (26.3) |
| History of hypertension | 292 (15.5) | 213 (33.9) |
| History of cardiovascular disease (angina, myocardial infarction or stroke) | 91 (4.8) | 54 (8.6) |

Unless otherwise specified, all values are reported as *n* (%)
[a]Quartiles of waist circumference in cm:
1, Females 52.7-70.5; Males 62.0-85.9
2, Females 70.6-76.9; Males 86.0-91.9
3, Females 77.0-85.4; Males 92.0-98.3
4, Females 85.5-137.0; Males 98.4-131.0
[b]Quartiles of total energy intake in kJ/day:
1, Females 3,214-6,278; Males 3,755-7,369
2, Females 6,279-8,006; Males 7,370-9,213
3, Females 8,007-9,845; Males 9,214-11,476
4, Females 9,846-18,831; Males 11,477-21,650

**Table 3: Odds ratios (ORs) and 95% confidence intervals (CIs) for the risk of type 2 diabetes by concentrations of 25(OH)D.**

| | | 25(OH)D[a] | | | | |
|---|---|---|---|---|---|---|
| | | **Median (IQR), nmol/L** | ***N*** | **Cases** | **OR[b] (95% CI)** | $p_{trend}$ |
| **Overall** | | | | | | |
| Quartiles | Q1 | 31.4 (7.3) | 623 | 200 | 1.00 (ref) | <0.001 |
| | Q2 | 42.9 (10.4) | 642 | 175 | 0.87 (0.65, 1.15) | |
| | Q3 | 53.0 (13.6) | 626 | 135 | 0.70 (0.52, 0.94) | |
| | Q4 | 72.1 (19.4) | 621 | 118 | 0.60 (0.44, 0.81) | |
| Per 25 nmol/L increase | | 48.3 (23.8) | 2,512 | 628 | 0.76 (0.63, 0.92) | 0.004 |
| **Wave 2** | | | | | | |
| Quartiles | Q1 | 30.8 (7.4) | 161 | 51 | 1.00 (ref) | 0.04 |
| | Q2 | 42.8 (11.6) | 155 | 43 | 0.73 (0.42, 1.30) | |
| | Q3 | 53.0 (13.1) | 159 | 38 | 0.78 (0.44, 1.38) | |
| | Q4 | 70.6 (20.5) | 165 | 28 | 0.51 (0.28, 0.93) | |
| Per 25 nmol/L increase | | 49.5 (23.2) | 640 | 160 | 0.68 (0.50, 0.93) | 0.02 |
| **Wave 3** | | | | | | |
| Quartiles | Q1 | 31.7 (7.1) | 462 | 149 | 1.00 (ref) | 0.01 |
| | Q2 | 42.9 (9.9) | 487 | 132 | 0.92 (0.66, 1.27) | |
| | Q3 | 52.8 (13.9) | 467 | 97 | 0.68 (0.49, 0.96) | |
| | Q4 | 72.9 (18.9) | 456 | 90 | 0.64 (0.45, 0.91) | |
| Per 25 nmol/L increase | | 48.1 (24.0) | 1,872 | 468 | 0.79 (0.64, 0.97) | 0.03 |

CI = confidence interval; IQR = interquartile range; OR = odds ratio; Q = quartile; *N* = number of participants; ref = reference

[a]Plasma equivalent concentrations adjusted for batch and seasonal effects.

[b]Adjusted for sex, age, country of birth, socioeconomic status, education, alcohol, smoking, physical activity, waist circumference, Mediterranean diet score, margarine intake, total energy intake, history of hypertension at baseline, and history of cardiovascular disease at baseline.

**Figure legends**

**Figure 1: Flow diagram of participants included in the main analyses.**
Participants included in the main analyses are shown in shaded boxes. A random sample of 480 participants was selected as wave 2 controls from the 2,391 subcohort participants without diabetes at wave 2. There were 1,491 subcohort participants who were not selected as wave 2 controls who attended wave 3, and did not have missing data on diabetes, and did not develop diabetes before wave 3. From these participants, a random sample of 1,404 participants was selected as wave 3 controls. 'Dead' at wave 2 means the participant did not complete the questionnaire and died between baseline and 20 June 2002 (final date of wave 2). 'Dead' at wave 3 means the participant did not attend wave 3 and died between wave 2 and 16 June 2007 (final date of wave 3).

**Figure 2: Odds ratios for 25(OH)D concentration and the incidence of type 2 diabetes.**
Odds ratios (ORs) are shown by quartiles of 25(OH)D and from analysis of restricted cubic splines (dashed curve).