# Supplementary material

# Vitamin D supplement use and associated demographic, dietary and lifestyle factors in 8024 South Asians aged 40–69 years: analysis of the UK Biobank cohort

# **Supplemental File 1**

### *Definition of vitamin D containing supplement use*

The Biobank baseline assessment did not ask participants to mark vitamin D and calcium supplements specifically but participants were able to choose both calcium and vitamin D as separate choices, so it is assumed that persons on combined calcium and vitamin D supplements would have marked that they used a vitamin D supplement. Similarly, the baseline assessment did not ask about dosage, brand or consistency of supplement use, so we were not able to take this in consideration in our analyses. It was not possible to analyse usage of cod liver oil as the only data available were on total fish oil supplementation usage (i.e. it was not possible to differentiate omega 3 fish oil supplement usage from that of cod liver oil).

### *Coding of non-dietary variables*

All variables were self-reported, being completed via a questionnaire on a touchscreen computer (ethnicity, gross household income, age completed education) at the baseline visit, via a questionnaire at the assessment centre reception (age at assessment visit, assessment centre attended), via a face to face interview at the baseline assessment (occupation) or derived from information provided before the visit (sex, Townsend Deprivation). Body Mass Index (BMI) was calculated from body weight and height when the participant underwent body composition assessment by bio-impedance at the baseline visit.

In terms of continuous variables, age when attended assessment centre was coded into 2 categories (<60 and ≥ 60 years old) to assess the impact of older vs. middle-age on supplement use. BMI was coded into a categorical variable according to usual clinical cut-offs: ≤25.4 kg/m2 Normal/Underweight; 26-29.4 kg/m2 overweight and 30 kg/m2 or higher obese, in order to make the results clinically meaningful. Underweight (BMI ≤17.4 kg/m2) was not classified as a separate category as only 16 (0.2%) of 7788 participants met this definition. Age completed education was retained as a continuous variable as it was not deemed relevant to categorise this. The Townsend Deprivation Index was derived from the participants’ postcodes by the Biobank investigators and was categorised for the current analysis as ≤ -1.05 (equal to or less than the UK median) and > -1.05 (above UK median). This was to make the interpretation of data more meaningful in terms of the UK average deprivation levels. Finally, daily time spent outdoors in summer was coded from a continuous variable to ≤30 minutes per day vs. >30 minutes per day, in order to provide meaningful exposures in relation to vitamin D production.

In terms of originally categorical variables, the gross household income variable was recoded from the initial 5 categories coded in the data (<£18K, £18-30.9K, £31-51.9K, £52-100K, >£100K) to 4 categories ((<£18K, £18-30.9K, £31-51.9K, ≥ 52K). This was because the final category (>100K) had only a relatively small number of participants (n=359) compared with the others. Assessment centre attended, a proxy for general area of residence, was recoded according to the 9 official regions of England (excluding the East of England where there were no assessment centres). The English region categories were then collapsed into 4 categories: Northern regions (North-West, North-East, Yorkshire and the Humber); Southern Regions (South-East and South-West); Greater London and The Midlands (East Midlands and West Midlands) to make comparisons meaningful. Scottish and Welsh centres were coded as Scotland and Wales respectively. Oily fish consumption, recorded via the touchscreen questionnaire, was originally coded as ‘never, less than once per week, once per week, 2-4 times per week, 5-6 times per week and once or more daily’. This was reclassified as ‘never, less than once per week, once per week and 2 or more times per week’ due to relatively low subject numbers in the highest two categories.

Ethnicity was not recoded, being retained as Bangladeshi, Indian and Pakistani. Occupational category was coded by UK Biobank investigators in accordance with the Standard Occupational Classification 20001. This classification includes the following categories: managers and senior officials (e.g. local government officials, chief executives); professional occupations (e.g. health professionals, accountants, scientists, teachers, legal professionals); associate professional and technical occupations (e.g. police, nurses, Information Technology workers, therapists); administrative and secretarial professions; skilled trades occupations (e.g. construction, agriculture); personal service occupations (e.g. childcare, hairdressing, travel services); sales and customer service occupations; process, plant and machine operatives (e.g. taxi drivers, assembly line operatives) and elementary occupations (e.g. farm workers, cleaners, security guards, building labourer). These main groupings were retained as the categories to be used in the analyses as no interpretive benefit was thought to be likely if job categories were made broader or narrower.

# **Supplemental File 2**

## *Participant characteristics*

For occupation, there was an association with sex (P<0.001), with over twice as many women than men in administrative and secretarial professions (21% vs. 8%), four times as many women than men in personal service occupations (9% vs. 2%) and one-tenth as many women than men in skilled trades occupations (1% vs. 10%). Men were more likely to be smokers than females (3% vs 14%, P<0.001). There were little relevant differences in the other variables (age, BMI, Townsend Deprivation Index, being born outside of the UK and the Republic of Ireland, decade of immigration, gross household income, assessment centre attended and geographical region, self-reported health status). Median (IQR) for days per week spent walking was similar in males and females (5(4) in both groups; P<0.001; Mann-Whitney U Test) but age left education was 14(8) in females and 16(8) in males (P<0.001), suggesting males left school slightly later than did females.

There was an association between ethnicity and self-reporting of health (P<0.001). Over half of Bangladeshis (59%) and Pakistanis (52%) reported only ‘fair or poor health’ rather than ‘excellent’ or ‘good’ health, compared with 39% of Indians. For women, there was an association between post-menopausal status and ethnicity (P<0.001) with only 50% of Pakistani women being postmenopausal, compared with 57% of Bangladeshi and 64% of Indian women.

There was a weak association between percentage of persons born outside the UK and the Republic of Ireland, with 88% of Pakistanis, 90% of Indians and 95% of Bangladeshis having immigrated to the UK. A higher proportion of Indians entered the UK in the 1960s/70s (73%, P<0.001) than that of Pakistanis (60%) and Bangladeshis (46%) and more Bangladeshis entered the UK in the 1980s/90s (44%), compared with 27% of Pakistanis and only 17% of Indians. There was an association between gross household income and ethnicity (P<0.001), with 2.3 times more Bangladeshis than Indians, and 1.2 times more Bangladeshis than Pakistanis having a gross household income of <£18K. Accordingly, in the top income bracket (>100K), there were more Indians (7%) as compared with Pakistanis (4%) and Bangladeshis (1%). Twice as many Indians than Bangladeshis and 1.5 times more Indians than Pakistanis lived in areas less than or equal to the UK median deprivation (i.e. less deprived neighbourhoods).

There was an association between assessment centre attended and ethnicity (P<0.001), with over two-thirds of Indians coming from the top 4 centres for South Asian recruitment (Leeds, Hounslow, Croydon and Birmingham) compared with around half of Pakistanis and Bangladeshis. Consequently, there was also an association between region and ethnicity (P<0.001), with around 50% of Bangladeshis and Indians coming from Greater London, compared with 24% of Pakistanis, who were more prevalent in the North (41%) than the other two groups. Similar proportions of each ethnic group came from the Midlands and South regions as well as Scotland and Wales. For occupation, there was an association with ethnicity (P<0.001), with just under half as many Bangladeshis in professional occupations as were Indians and Pakistanis, but with around twice as many Bangladeshis in skilled trades and elementary occupations than the other two groups. Also, there were twice as many Pakistanis as process, plant and machine operatives than Indians and Bangladeshis. Indians had a higher age of completion of education (16(7) as compared with Bangladeshis (11(10)) and Pakistanis (13(9)) (P<0.001). Indians had a higher number of days per week walking for 10 minutes or more (median (IQR) 5(2)) compared with the other two groups (4(3))(P<0.001, Kruskal-Wallis Test).

## *Logistic regression models*

When the dietary variables (vegetarianism, oily fish consumption) were initially added to model 4, the Hosmer and Lemeshow Test had a P value of 0.04, suggesting the data did not fit the model well. Removal of vegetarianism gave a better model fit (Hosmer and Lemeshow Test=0.43) than did removal of oily fish intake (Hosmer and Lemeshow Test=0.07), so oily fish intake was retained in the model rather than vegetarianism.

In the final model (model 4), the relative contributions to the pseudo-R2 value (Nagelkerke) for each variable were as follows: sex (61%); region (10%); household income (8%); ethnicity (7%); age (7%); BMI (3%), oily fish intake (3%). Inspection of collinearity diagnostics highlighted no likely issues were present (Tolerance= 0.95 to 0.98, VIF= 1.02 to 1.11, depending on variable), and there were only three outliers (standardised Z residuals greater than 2). The model correctly predicted 99% of those not on a supplement, but only 3% of those actually on a supplement.

Sub-analyses were also conducted on model 4 to assess whether assessing single vitamin D supplements separately from multivitamin supplements influenced the results (participants who used both were excluded from these analyses). See supplementary table 1 for specific details of these results.

For single vitamin D supplementation, sex was still an associated variable (OR=2.41 95%CI 1.94 to 3.00). Age and region were still associated variables, with effect sizes little changed from before. The oily fish category different from the reference category (twice or more per week) was now ‘less than once per week’, rather than ‘never but the upper confidence limit’. For multivitamin and mineral use, sex was still an associated variable but the effect size was slightly attenuated (OR=1.77 95% CI 1.56 to 2.00). Age, region and oily fish consumption were now not associated variables, and gross household income showed a slight attenuation in the effect size.

# **Supplemental File 3**

*Completion rates of questions*

Although the majority of the other questions had excellent completion rates (>95%), some of the variables had a relatively large amount of missing data with completion rates as follows: job type (70%); age completed education (61%) and gross household income (74%). Income and age completed education were completed by touchscreen and face to face interview respectively at the baseline visit, which generally had very high compliance with all participants answering nearly all the questions. The lack of completion suggests that the participants did not know the answers to these particular questions or preferred not to answer them. For age completed education, 4.1% of the n 8024 participants preferred not to answer, 1.7% had never gone to school and 4.0% did not know. Another 40% of participants had no recorded response, which could have potentially biased the results. For gross household income 17% preferred not to answer, 8.3% did not know, and 1.2% gave no response. Not knowing gross household income may have led to a loss of these people from the logistic regression model which could be a source of bias.

# **Supplementary Tables**

## **Supplemental Table 1: Relationship between potential associated variables and use of vitamin D containing supplements**

|  | **Supplement Use\*** | | | |  | **χ2 test**† |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **% Yes** | **n** | **% No** | **n** | **Total n** | **P value** | **Result** |
| **Sex** |  |  |  |  |  |  |  |
| Female | 38.6 | 1393 | 61.4 | 2218 | 3611 | <0.001 | Included |
| Male | 22.6 | 935 | 77.4 | 3207 | 4142 |
| **Ethnicity** |  |  |  |  |  |  |  |
| Bangladeshi | 22.5 | 49 | 77.5 | 169 | 218 | <0.001 | Included |
| Indian | 31.9 | 1849 | 68.1 | 3944 | 5793 |
| Pakistani | 24.7 | 430 | 75.3 | 1312 | 1742 |
| **Age Group** | | | | | | |  |
| 40-59 years | 28.7 | 1609 | 71.3 | 3994 | 5603 | <0.001 | Included |
| 60 years and over | 33.4 | 719 | 66.6 | 1431 | 2150 |
| **BMI kg/m2** | | | | | | |  |
| ≤25.4 Normal/Underweight | 32.8 | 897 | 67.2 | 1837 | 2734 | 0.001 | Included |
| 26-29.4 Overweight | 28.8 | 811 | 71.2 | 2007 | 2818 |
| 30 or higher Obesity | 28.5 | 566 | 71.5 | 1420 | 1986 |
| **Gross household income:** | | | | | | |  |
| <£18 000 | 25.8 | 476 | 74.2 | 1369 | 1845 | <0.001 | Included |
| £18 000 to £30 900 | 30.6 | 423 | 69.4 | 958 | 1381 |
| £31 000 to £51 900 | 33.0 | 389 | 67.0 | 790 | 1179 |
| ≥£52 000 | 29.9 | 416 | 70.1 | 977 | 1393 |
| **Townsend Deprivation Index** | | | | | | | |
| ≤ UK Median‡ | 30.7 | 899 | 69.3 | 2033 | 2932 | 0.33 | Not Included |
| >UK Median‡ | 29.6 | 1425 | 70.4 | 3387 | 4812 |  |
| **Region** | | | | | | | |
| Northwest, North-East, Yorkshire and the Humber | 25.6 | 454 | 74.4 | 1320 | 1774 | <0.001 | Included |
| South East, South West | 28.1 | 174 | 71.9 | 445 | 619 |  |  |
| Wales | 18.1 | 21 | 81.9 | 95 | 116 |  |  |
| Scotland | 23.6 | 42 | 76.4 | 136 | 178 |  |  |
| East Midlands, West Midlands | 26.0 | 412 | 74.0 | 1174 | 1586 |  |  |
| Greater London | 35.2 | 1225 | 64.8 | 2255 | 3480 |  |  |
| **Vegetarian** | | | | | | | |
| No | 28.9 | 1504 | 71.1 | 3696 | 5200 | <0.001 | Trialled |
| Yes | 35.7 | 517 | 64.3 | 931 | 1448 |  |  |
| **Oily Fish Consumption** |  |  |  |  |  |  |  |
| Never | 31.9 | 807 | 68.1 | 1724 | 2531 | 0.002 | Included |
| <Once per week | 27.1 | 582 | 72.9 | 1567 | 2149 |  |  |
| Once per week | 31.4 | 631 | 68.6 | 1381 | 2012 |  |  |
| Twice or more per week | 31.6 | 259 | 68.4 | 560 | 819 |  |  |
| **Daily Time outdoors in summer** | | | | | | | |
| ≤ 30 minutes | 32.1 | 238 | 67.9 | 504 | 742 | 0.20 | Not included |
| >30 minutes | 29.8 | 1879 | 70.2 | 4433 | 6312 |  |  |

\* Either a single vitamin D supplement, a multi-vitamin and mineral supplement or both, † alpha=0.005 used due to Bonferroni correction of multiple testing ‡ -1.1 is UK median (<https://census.ukdataservice.ac.uk/get-data/related/deprivation>)

## **Supplemental Table 2: Characteristics of n 8024 South Asian UK Biobank Participants by Sex: Continuous Data**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Female n 3730** | | | | | **Male n 4294** | | | | |  |
|  | **Mean** | **SD** | **n** | **Lower 95%CI** | **Upper**  **95%CI** | **Mean** | **SD** | **n** | **Lower 95%CI** | **Upper**  **95%CI** | **P\*** |
| **Age years** | 53 | 8 | 3730 | 53 | 53 | 54 | 9 | 4294 | 53 | 54 | 0.01 |
| **BMI kg/m2** | 28 | 5 | 3674 | 27 | 28 | 27 | 4 | 4114 | 27 | 27 | <0.001 |
| **Waist: Hip ratio** | 0.8 | 0.1 | 3680 | 0.8 | 0.9 | 1.0 | 0.1 | 4227 | 1.0 | 1.0 | <0.001 |
|  | **Median** | **IQR** | **n** |  |  | **Median** | **IQR** | **n** |  |  |  |
| **Vitamin D intake (micrograms/ day)** | 1.0 | 1.6 | 1023 |  |  | 1.2 | 2.0 | 1183 |  |  | 0.002 |

CI, confidence interval. SD, standard deviation. \*One way ANOVA; except Kruskal Wallis Test (with Dunn’s multiple comparison post hoc tests) for vitamin D intake. Like superscripts within rows depict statistically significant differences highlighted in post-hoc tests (P<0.05).

## **Supplemental Table 3: Characteristics of n 8024 South Asian UK Biobank Participants by Sex: Categorical data**

|  | **Female n 3730** | | **Male n 4294** | |  |
| --- | --- | --- | --- | --- | --- |
|  | **%** | **n** | **%** | **n** | **P\*** |
| **Ethnicity:** | | | | | |
| **Bangladeshi** | 2 | 74 | 4 | 162 | <0.001 |
| **Indian** | 79 | 2939 | 70 | 3012 |
| **Pakistani** | 19 | 717 | 26 | 1120 |
| **Current smoker (% Y; any frequency)** | 3 | 3720 | 14 | 4279 | <0.001 |
| **Oily fish intake** | | | | | |
| **Never** | 41 | 1476 | 27 | 1125 | <0.001 |
| **<Once per week** | 25 | 902 | 32 | 1307 |
| **Once per week** | 25 | 897 | 29 | 1189 |
| **2-4 times per week** | 8 | 296 | 10 | 430 |
| **5-6 times per week** | 1 | 29 | 1 | 45 |
| **Once or more daily** | 0.4 | 15 | 0.8 | 31 |
| **Vegetarian %Y** | 29 | 959 | 15 | 525 | <0.001 |
| **% Reporting Fair/Poor Health** | 44 | 3660 | 42 | 4227 | 0.15 |
| **Post-menopausal (% Y)\*\*** | 55 | 3651 | - | - | - |
| **Born outside of UK and Republic of Ireland (% Y)** | 88 | 3216 | 90 | 3979 | 0.003 |
| **Of which:** | | | | | |
| **% Immigrated Before 1959** | 2 | 79 | 2 | 78 | <0.001 |
| **% Immigrated 1960-1979** | 71 | 2232 | 68 | 2572 |
| **% Immigrated 1980-1999** | 20 | 643 | 19 | 729 |
| **% Immigrated 2000 onwards** | 7 | 207 | 10 | 385 |
| **Gross household income: (£)** | | | | | |
| **<18 000** | 33 | 825 | 32 | 1087 | 0.04 |
| **18 000-30 900** | 24 | 614 | 23 | 789 |
| **21 000-51 900** | 21 | 526 | 20 | 663 |
| **52 000-100 000** | 17 | 427 | 18 | 614 |
| **>100 000** | 5 | 129 | 7 | 230 |
|  |  |  |  |  |  |
| **Townsend Deprivation Index** | | | | | |
| **≤ UK median (Less deprived)** | 39 | 1448 | 36 | 1558 | 0.02 |
| **> UK median (More deprived)** | 61 | 2277 | 64 | 2732 |  |
| **Biobank Assessment Centre** | | | | | |
| **Leeds** | 8 | 290 | 9 | 372 | <0.001 |
| **Hounslow** | 33 | 1228 | 28 | 1215 |
| **Croydon** | 13 | 479 | 11 | 456 |
| **Birmingham** | 15 | 550 | 16 | 701 |
| **Other** | 32 | 1183 | 36 | 1550 |
| **Region** | | | | | |
| **Northern England** | 21 | 789 | 26 | 1093 | <0.001 |
| **Southern England** | 8 | 311 | 7 | 319 |
| **Wales** | 1 | 49 | 2 | 67 |
| **Scotland** | 2 | 77 | 3 | 108 |
| **The Midlands** | 19 | 721 | 22 | 924 |
| **Greater London** | 48 | 1783 | 42 | 1783 |

Y, Yes. \* based on Chi-Square Test

**Supplemental Table 4: Odds of being a vitamin D containing supplement user\*: Either single vitamin D supplement, or as part of multivitamin and mineral supplement)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model 4 All participants** | **Single vitamin D supplement n 5512** | | | | | | **Multivitamin and Mineral Supplement n 5512** | | | | | |
|  | **n** | **B\*** | **SE** | **OR**† | **Lower 95% CI** | **Upper 95% CI** | **n** | **B** | **SE** | **OR**† | **Lower 95% CI** | **Upper 95% CI** |
| **Sex** |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | 2403 | 0.88 | 0.11 | 2.41 | 1.94 | 3.00 | 2403 | 0.57 | 0.06 | 1.77 | 1.56 | 2.00 |
| Male | 3109 |  |  | 1.00 |  |  | 3109 |  |  | 1.00 |  |  |
| **Ethnicity** |  |  |  |  |  |  |  |  |  |  |  |  |
| Indian | 4177 |  |  | 1.00 |  |  | 4177 |  |  | 1.00 |  |  |
| Pakistani | 1196 | 0.37 | 0.14 | 1.45 | 1.11 | 1.90 | 1196 | -0.15 | 0.09 | 0.86 | 0.73 | 1.02 |
| Bangladeshi | 139 | -0.68 | 0.47 | 0.51 | 0.20 | 1.28 | 139 | -0.33 | 0.23 | 0.72 | 0.46 | 1.13 |
| **Body Mass Index**‡ |  |  |  |  |  |  |  |  |  |  |  |  |
| ≤25.4 NormalNormal/Underweight | 2055 | 0.19 | 0.14 | 1.21 | 0.92 | 1.57 | 2055 | 0.20 | 0.08 | 1.22 | 1.04 | 1.43 |
| 26-29.4 Overweight | 2041 | 0.04 | 0.14 | 1.04 | 0.79 | 1.36 | 2041 | 0.02 | 0.08 | 1.02 | 0.86 | 1.20 |
| 30 or higher Obesity | 1416 |  |  | 1.00 |  |  | 1416 |  |  | 1.00 |  |  |
| **Age group** |  |  |  |  |  |  |  |  |  |  |  |  |
| 40-59 years | 4076 | -0.30 | 0.12 | 0.74 | 0.59 | 0.93 | 4076 | -0.18 | 0.07 | 0.83 | 0.72 | 0.96 |
| ≥60 years | 1436 |  |  | 1.00 |  |  | 1436 |  |  |  |  |  |
| **Household Income**§ |  |  |  |  |  |  |  |  |  |  |  |  |
| <18K | 1711 | 0.26 | 0.15 | 1.30 | 0.96 | 1.75 | 1711 | -0.36 | 0.09 | 0.70 | 0.59 | 0.84 |
| 18 to 30.9K | 1314 | 0.08 | 0.16 | 1.08 | 0.79 | 1.48 | 1314 | -0.02 | 0.09 | 0.98 | 0.82 | 1.17 |
| 31 to 51.9K | 1134 | 0.04 | 0.17 | 1.04 | 0.75 | 1.44 | 1134 | 0.12 | 0.09 | 1.13 | 0.94 | 1.35 |
| ≥52K | 1353 |  |  | 1.00 |  |  | 1353 |  |  | 1.00 |  |  |
| **Region** |  |  |  |  |  |  |  |  |  |  |  |  |
| North England | 1305 | -0.57 | 0.15 | 0.57 | 0.43 | 0.75 | 1305 | -0.17 | 0.09 | 0.84 | 0.71 | 0.99 |
| South England | 487 | -0.66 | 0.22 | 0.52 | 0.33 | 0.79 | 487 | -0.16 | 0.12 | 0.86 | 0.68 | 1.07 |
| Wales | 85 | -1.45 | 0.72 | 0.24 | 0.06 | 0.97 | 85 | -0.43 | 0.29 | 0.65 | 0.37 | 1.14 |
| Scotland | 133 | -0.85 | 0.43 | 0.43 | 0.18 | 0.99 | 133 | -0.47 | 0.23 | 0.62 | 0.39 | 0.99 |
| Midlands | 1119 | -0.62 | 0.15 | 0.54 | 0.40 | 0.73 | 1119 | -0.18 | 0.09 | 0.84 | 0.71 | 0.99 |
| Greater London | 2383 |  |  | 1.00 |  |  | 2383 |  |  | 1.00 |  |  |
| **Oily Fish Consumption** |  |  |  |  |  |  |  |  |  |  |  |  |
| Never | 1672 | -0.19 | 0.18 | 0.83 | 0.58 | 1.18 | 1672 | -0.14 | 0.11 | 0.87 | 0.70 | 1.08 |
| <Once per week | 1713 | -0.50 | 0.19 | 0.61 | 0.42 | 0.87 | 1713 | -0.19 | 0.11 | 0.83 | 0.67 | 1.03 |
| Once per week | 1519 | -0.10 | 0.18 | 0.91 | 0.64 | 1.29 | 1519 | -0.02 | 0.11 | 0.99 | 0.79 | 1.23 |
| 2 or more times per week | 608 |  |  | 1.00 |  |  | 608 |  |  | 1.00 |  |  |
| **Constant** |  | -2.54 | 0.24 | 0.08 |  |  |  | -0.98 | 0.14 | 0.38 |  |  |

SE, Standard Error. \*B=unstandardised coefficient †OR= odds of being a supplement user (non-supplement user OR=1); ‡kg/m2; §gross household income.

## **Reference**

1. ONS (2000). Standard Occupational Classification Vol 1. Structure and descriptions of unit groups. London:TSO