Predicting ambient ultraviolet from routine meteorological data; its potential use as an instrumental variable for vitamin D status in pregnancy in a longitudinal birth cohort in the UK

Adrian Sayers,1* Kate Tilling,2 Barbara J Boucher,3 Kate Noonan4 and Jon H Tobias1

Accepted 26 May 2009

Background Maternal vitamin D status in pregnancy has been postulated to have important effects on intrauterine development. UVB radiation is not commonly measured but is the prime determinant of circulating 25-hydroxyvitamin-D [25-(OH)D] and is highly dependent on regional weather including cloud cover, ozone and sunshine hours.

Methods Using linear regression we described the relationship between estimated ambient-erythemal ultraviolet (eUV) exposure in Oxford (1990–95) and total hours of sunshine and month in order to forecast eUV in nearby regions, whilst adjusting for regional variations in weather. The forecast was validated with empirical data collected from Cornwall and then predicted for the Avon region. Total 98-day prenatal ambient-eUV was then predicted in 355 expectant mothers in the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort and its relationship with maternal vitamin D status was determined.

Results Estimated ambient-eUV was strongly associated with measured ambient-eUV \( (r^2 = 0.989) \) with a near 1:1 prediction for the validation data set \( (\beta = 0.99, 95\% \text{ confidence interval (CI)} 0.913, 1.067 \) \( r^2 = 0.980 \) \); strong seasonal associations were observed between eUV in the last trimester of pregnancy and maternal serum 25-(OH)D concentrations \( (r^2 = 0.40) \).

Conclusion This technique of prediction could be applied to existing cohorts allowing the relationship between maternal vitamin D status and the health of the offspring to be studied via instrumental variable analysis.

Keywords Epidemiology, maternal exposure, pregnancy, prenatal exposure delayed effects, seasons, ultraviolet rays, vitamin D, instrumental variable, ALSPAC

---

1 Academic Rheumatology, Department of Clinical Science at North Bristol, University of Bristol, Bristol, UK.
2 Department of Social Medicine, University of Bristol, Bristol, UK.
3 Centre for Diabetes and Metabolic Medicine, Bart’s and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK.
4 Department of Clinical Chemistry, Bart’s and The Royal London NHS Trust, London, UK.
* Corresponding author. Academic Rheumatology, Avon Orthopaedic Centre, Southmead Hospital, Bristol BS10 5NB, UK. E-mail: adrian.sayers@bristol.ac.uk
Introduction

Vitamin D is an important pro-hormone and vitamin D insufficiency has been implicated in a number of different non-communicable diseases, including osteoporosis,1,2 coronary heart disease,3,4 peripheral arterial disease,5 colon cancer,6 prostate cancer,7 breast cancer,8 rheumatoid arthritis9 and type I diabetes.10 However, it is not clear for how long, or at what stage of life an individual needs to be vitamin D insufficient or deficient, in order to become predisposed to an increased risk of various diseases.

There is some evidence to suggest that in utero exposure to vitamin D is important in determining long-term outcomes including bone health,11,12 prostate growth,13 diabetes14 and asthma.15 However, the typical methods used to study the effects of vitamin D status on health outcomes are either observational,16 or from an ecological perspective.6,10 A common problem of observational or ecological data intrinsic to the method of collection is confounding and the ‘ecological fallacy’, respectively. However, it is difficult to identify the extent of this problem without employing more robust methods.

Although there are no randomized control trials that report long-term health outcomes of vitamin D supplementation in pregnancy, several trials are underway in this area e.g. Maternal Vitamin-D Osteoporosis Study (MAVIDOS) trial in Southampton, UK.17 However, randomized control trials into the effects of vitamin D intakes on health outcomes can suffer from treatment contamination (confounding) by increased exposure to naturally occurring UVB. Whilst it may be possible to control for the differing effects of exposure by UVB-sensitive badges,18 it increases the burden for the participant and may introduce another form of bias.

An alternative method of investigating causal effects is via instrumental variable analysis.19,20 In order for a variable to be considered an ‘instrument’, it must be (i) independent of measured or unmeasured confounders (e.g. ambient-UVB radiation is unrelated to social economic position or any other confounder); (ii) associated with the exposure of interest (e.g. ambient-UVB radiation is associated with vitamin D status, given that UVB exposure is the prime determinant of vitamin D); and conditionally independent of the health outcome given the exposure and the measured or unmeasured confounders (e.g. ambient-UVB is unrelated to the health outcome after taking into account vitamin D status and the confounders). Because of the requirements of an instrument, the effect of the instrument (ambient-UVB) on the health outcome of interest (e.g. bone mass) is the product of the effect of the instrument (ambient-UVB) on exposure (vitamin D status), and the effect of the exposure (vitamin D status) on the health outcome of interest, therefore providing an unbiased estimate of the effect of the exposure (vitamin D status) independent of confounders.19 We are unaware of any previous studies applying this technique of instrumental variable analysis to the study of vitamin D status.

25-hydroxyvitamin-D [25-(OH)D] is used as a marker of vitamin D status, and routine assays often assess both 25-(OH)D3 (cholecalciferol) and 25-(OH)D2 (ergocalciferol). Vitamin D3 is primarily synthesized from exposure to UVB radiation (280–315 nm), with small quantities also contained in oily fish and cod liver oil, whereas vitamin D2 is consumed within the diet. Exposure to UVB is dependent on intrinsic factors (e.g. skin type, age, clothing/skin protection,21 personal behaviour22) and extrinsic factors (e.g. solar zenith angle (reflecting season, time of day and latitude), ozone, cloud cover, pollution and surface reflection). Knowledge of extrinsic factors can be generalized to a population, whereas knowledge of intrinsic factors relies on accurate self-reporting or compliance with UVB-exposure meters, both of which are subject to confounding and other biases, and can be costly to collect.

Exploring the role of vitamin D on the intrauterine environment is impeded by the lack of availability of study populations in which vitamin D status or UVB exposure in the last trimester of pregnancy has been measured. However, since date and place of birth are widely available in nearly all cohorts, it should be possible to estimate ambient-UVB in pregnancy by combining this information with meteorological data, and subsequently make inferences with regards to the effects of UVB which is well known to be the primary source of vitamin D.23–27

We describe a method of estimating ambient-erythemal ultraviolet (eUV) from total hours of sunshine and month of year, which we then validated with data from weather stations of similar latitude and altitude. Subsequently, we confirmed that cumulative ambient-eUV exposure in the last trimester of pregnancy provides a useful estimate of maternal vitamin D status in the last trimester of pregnancy, based on measured concentrations of serum [25-(OH)D] in 355 expectant mothers from the Avon Longitudinal Study of Parents and Children (ALSPAC). We propose that ambient-eUV, derived using the method reported here, can be used to investigate the effect of maternal vitamin D status on a variety of different outcomes through the associations described in Figure 1.

![Figure 1 Proposed directed acyclic graph linking ambient-UVB, UVB exposure, cutaneous vitamin D synthesis and health outcomes](image-url)
Materials and methods

Participants
ALSPAC is a geographically-based birth cohort study, investigating factors influencing the health, growth and development of children. All pregnant women resident within a defined part of the former county of Avon in South West England with an expected date of delivery between April 1991 and December 1992 were eligible for recruitment, of whom less than 14 000 were enrolled28 (http://www.alspac.bristol.ac.uk). Ethical approval was obtained from the ALSPAC Law and Ethics committee and relevant local ethics committees. Data in ALSPAC are collected by self-completion postal questionnaires sent to parents, by linkage to computerized records, by abstraction from medical records and from examination of the children at research clinics.

[25-(OH)D] determination
[25-(OH)D] was measured in serum from blood samples in 355 pregnant mothers using a chemiluminescence immuno-assay technique (DIASORIN 13040 Analyser, Saluggia, Italy) recognizing both 25-(OH)D2 and 25-(OH)D3 within and between batch precision for low and high QC (2005–06), 10–12 and 12–15%, respectively. Assay performance was within external-quality-scheme standards (DEQUAS) during this period.

eUV and meteorological data
The National Radiological Protection Board (NRPB), now part of the Health Protection Agency, recorded eUV at a number of different sites around the UK at the time mothers were being recruited (1990–94). eUV is a measure of UV exposure, which weights the wavelength according to its harmful effects (erythema). eUV is reported to the general population as it reflects how long you can stay in the sun (sun burn index). High levels of eUV are principally composed of short wavelengths of UVB, thus eUV is a reasonable proxy for UVB exposure. This relationship has been shown to hold with an error of <10% for all solar zenith angles (the position of the sun in the sky, which reflects season, time of day and latitude), where UVB with a wavelength of 280–315 nm is ~7.55 (eUV).29 Due to the proportionality between eUV and UVB, results will be presented with respect to eUV.

The NRPB measured eUV in Chilton, Oxfordshire (~60 miles East North East from Avon), from 1990, and Camborne, Cornwall (~180 miles South West from Avon), from 1993 (partially over the period of interest), but no eUV measurements were taken in the Avon region, which is geographically positioned between Oxford and Cornwall. Archive weather data, which included total hours of sunshine, were recorded at local Meteorological Office weather stations in Oxford, Cornwall and Avon from 1990.

Statistical analyses

Model generation
Due to the extrinsic properties that determine eUV exposure, specifically solar zenith angle and cloud cover, it is possible to derive a prediction model via linear regression using total monthly recorded sunshine (inversely related to cloud cover) and month (an indicator of average solar zenith angle) as a categorical exposure to predict eUV in Oxford. Model 1 assumes all other extrinsic factors such as ozone and surface reflection are constant. Higher-order sun terms were fitted (e.g. hours of sunshine squared and sunshine cubed), and nested models were compared using likelihood ratio tests to determine the most parsimonious model with the best fit. The fitted model may lead to an underestimation of residual variation as total monthly hours of sunshine are used instead of daily recorded sunshine. However, monthly estimates are simple to handle and are more commonly available than daily records.

Model validation
The validity of the model was tested using the eUV and hours of sunshine data collected in Cornwall. The model, derived as above, was used to predict eUV from the sunshine measures. These predicted eUV measures were then compared with the actual eUV measures recorded. This model assumes that any regional weather patterns are encapsulated in the local sunshine measurements.

Local eUV estimation
We then used this validated model to predict eUV exposure in Avon, adjusting for regional weather variations by using measures of local sunshine hours per day.

Individual last trimester prediction
Using locally estimated eUV, we estimated cumulative ambient-eUV in the 98 days pre-birth, i.e. in the last trimester of pregnancy, due to its postulated importance in neonatal bone development,30 for every mother in the cohort. Monthly totals of predicted ambient-eUV were calculated and proportionally assigned over the 98 days pre-birth. This method of estimation assumes that extrinsic factors that influence eUV are uniform across the region.

Cumulative ambient-eUV in the last trimester of pregnancy and maternal [25-(OH)D]
The association between eUV in the last trimester and measured vitamin D status was investigated using linear regression, evidence of homoscedasticity led to [25-(OH)D] being loge transformed and a non-linear model fitted by ordinary least squares (OLS).
Results

Model 1 described above proved to explain most of the residual variation between eUV, sunshine and month. However, the addition of total sunshine^2 (Model 2) resulted in a significant improvement in fit \( P < 0.0001, r^2 = 0.9889 \) and this model was therefore used for all analyses and predictions. Parameter estimates are listed in Table 1. Model 2 was then validated from data collected in Cornwall. There was no evidence of a mean difference between observed and predicted eUV (mean dif. (pred–obs) = 0.19, standard deviation (SD) = 0.87, [95% confidence interval (CI) (pred–obs): −0.25, 0.64]); 82.35% of predicted values were within 1 watts h m\(^{-2}\) eUV of observed values and 100% of values were within 2.5 watts h m\(^{-2}\) eUV of observed values. eUV varies seasonally from 2.4 watts h m\(^{-2}\) in the winter to 52.9 watts h m\(^{-2}\) in the summer. Pitman’s test confirmed that there was no evidence that the measured or estimated standard deviation were different (SD eUV measured = 6.2, SD eUV predicted = 6.2, \( P = 0.99 \)). The resulting predictions for Oxford, Cornwall and Avon are shown in the top, middle and lower panel of Figure 2, respectively. As has been shown, the model for predicting eUV from sunshine meteorological data in Oxford was able to predict eUV in Cornwall with a small level of error.

Table 1 Linear regression estimates used to estimate cumulative eUV exposure in the last trimester of pregnancy from data collected in Oxford, UK, from January 1990 to December 1994. January was coded as the baseline month

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.194</td>
<td>−0.32, 2.70</td>
</tr>
<tr>
<td>Monthly sunshine hours</td>
<td>−0.022</td>
<td>−0.05 to 0.01</td>
</tr>
<tr>
<td>Quadratic monthly sunshine hours</td>
<td>0.0002</td>
<td>0.0002 to 0.0002</td>
</tr>
</tbody>
</table>

Mean monthly increases in eUV^a compared with January

<table>
<thead>
<tr>
<th>Month</th>
<th>Increase (eUV)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>February</td>
<td>0.746</td>
<td>−0.31 to 1.80</td>
</tr>
<tr>
<td>March</td>
<td>2.957</td>
<td>1.67 to 4.25</td>
</tr>
<tr>
<td>April</td>
<td>5.478</td>
<td>3.93 to 7.03</td>
</tr>
<tr>
<td>May</td>
<td>9.549</td>
<td>7.92 to 11.18</td>
</tr>
<tr>
<td>June</td>
<td>12.942</td>
<td>11.33 to 14.56</td>
</tr>
<tr>
<td>July</td>
<td>13.029</td>
<td>11.33 to 14.73</td>
</tr>
<tr>
<td>August</td>
<td>9.606</td>
<td>7.88 to 11.33</td>
</tr>
<tr>
<td>September</td>
<td>6.120</td>
<td>4.68 to 7.56</td>
</tr>
<tr>
<td>October</td>
<td>2.218</td>
<td>0.92 to 3.52</td>
</tr>
<tr>
<td>November</td>
<td>0.449</td>
<td>−0.59 to 1.48</td>
</tr>
<tr>
<td>December</td>
<td>−0.050</td>
<td>−1.08 to 0.98</td>
</tr>
</tbody>
</table>

^a eUV is measured in watts h m\(^{-2}\).

Following the prediction of monthly eUV estimates, the date 98 days pre-birth was calculated for all individuals in the cohort, and the total eUV exposure for each individual obtained by imputation. The imputation resulted in typical seasonal variations, with the peaks and troughs of cumulative eUV exposure delayed approximately one-and-a-half months compared with the summer maximums and winter minimums.

Using linear regression, the association between vitamin D status [serum 25-(OH)D concentration] (the outcome) and imputed cumulative eUV exposure in the 3 months prior to birth (the exposure) was
investigated in the 355 study mothers. There was a strong linear association where a one-unit increase in eUV resulted in a 1.04 nmol$^{-1}$ increase in serum \[25\text{-(OH)}D\] (95% CI 0.88, 1.22), $r^2 = 0.311$. However, the model was not homoscedastic and a non-linear model was fitted by OLS regression and log$_e$ transforming \[25\text{-(OH)}D\]; this model was homoscedastic, and the fit was improved ($r^2 = 0.400$, $P < 0.0001$). The best prediction of serum \[25\text{-(OH)}D\] from eUV is the non-linear model presented in Figure 3. Further analyses controlling for the age of mother at birth and differing time of venopuncture did not affect these findings (as they were orthogonal with eUV exposure).

**Limitations**

The model described depends on a number of assumptions, which include month being used as an indicator of solar zenith angle (which controls for seasonal changes in eUV irradiation) and hours of sunshine being inversely related to cloud cover, thus allowing seasonal and local weather adjustment. Whilst both assumptions are plausible, there may be minor violations with respects to cloud cover, since light cloud cover may unduly lower the hours of total sunshine but have very little effect on ambient levels of eUV.$^{21}$ In addition, the model takes no account of variation in pollution, ozone, altitude, latitude or surface reflection. In spite of these omissions, the estimates of eUV in Cornwall, based on data collected in Oxford, which is >200 miles away, were very accurate. Since Avon is located 140 miles closer to Oxford than Cornwall, our model should perform as well, if not better for Avon, although we have no data to verify this assertion. In addition, our model for predicting last trimester eUV is the integral of a model estimating total monthly eUV from total hours of sunshine and sunshine,$^{2}$ and month as a categorical indicator of solar zenith angle; this may not be the most efficient or parsimonious method, but its simplicity makes the method the most accessible.
Finally we assume that eUV is a good proxy for UVB.21,29

**Applications of the prediction**

The method described has a number of possible applications in elucidating associations between UVB exposure and health outcomes, including its use as an instrument in instrumental variable analysis.

Epidemiological studies often suffer from confounding, and the ability to make unbiased inferences is important. There may be many mechanisms by which true UVB exposure and maternal vitamin D status may be confounded. For example, socioeconomic position may influence the diet of the mothers, and their access to safe outside areas where they can be exposed to UVB. Whilst it is possible to adjust for proxies of socio-economic position, there is always the concern that residual confounding may exist, which is why we primarily rely on the use of randomized control trials to indicate causal associations. Recently epidemiology has adopted the use of instrumental variables to make causal inferences.19,20

For a variable to be considered an instrument, it must be only related to the outcome of interest through the proposed causal path of instrument, exposure and outcome (see Figure 4).

Ambient-UVB radiation satisfies the criteria as an instrument because (i) ambient-UVB in pregnancy is associated with vitamin D status through the causal path of ambient-UVB, actual UVB exposure and vitamin D synthesis; (ii) time of conception and subsequently ambient-UVB levels in the last trimester of pregnancy is unrelated to all measured and unmeasured confounders; and (iii) ambient-UVB radiation is independent of the health outcome of interest given an individual’s vitamin D status and confounding factors. This assumes that there is no other functional link between ambient-UVB and the health outcome of interest. If ambient-UVB in the last trimester of pregnancy is an instrument of true maternal UVB exposure, and true UVB exposure is the prime determinant of vitamin Dα, then it may be possible to estimate the causal effect of vitamin Dα on a number of different health outcomes of the child (Figure 4).

Because randomized control trials are costly and must be prospective by design, it will be many years before causal evidence can be provided about the effects of maternal vitamin D on long-term health outcomes of the child using such an approach. However, as date and location of birth can be collected retrospectively, then, with the judicious collection of meteorological data, it should be possible to reconstruct ambient levels of UVB in many established cohorts. After the instrument has been constructed, it will be possible to detect causal associations between maternal vitamin D concentration in the last trimester of pregnancy and a number of different health outcomes. As ambient-eUV explains 40% of the variation of serum 25(OH)D, we consider this a strong instrument, especially in comparison with many genetic instruments, which may explain <5% of the total variation of the phenotype. Therefore, the potential for unravelling many of the methodological problems associated with studies of maternal vitamin D exposure is great. In addition, if recorded data on actual maternal vitamin D concentrations exist, it will be possible to estimate the magnitude of the causal effect of maternal vitamin D concentration on the health outcome of interest.

The generalizability of this method and using environmental instruments in other cohorts requires careful consideration. Geographic and topographic separation of the cohort is an important factor to consider. ALSPAC is a birth cohort with a small catchment area of (1340 km²) which is only a little larger than New York City (1200 km²) and is not divided by mountain ranges, which can isolate weather patterns that influence cloud cover and therefore UVB; this, in turn, yields homogenous weather patterns across the county. However, if the cohort catchment area is large, topographically divided into areas of low and high social economic position, and ambient-eUV significantly differs in both locations; then this may be one way that the instrument assumption may be violated. However, if the populations are heterogeneous, and the location of the individual is known, stratified analyses could be performed, which should appropriately adjusted for any regional differences and still

---

**Figure 4** Causal directed acyclic graph illustrating the generic use of instrumental variables, its current application in genetic epidemiology and its proposed use as an environmental instrument.
allow generalizable results. Therefore, careful consideration of cohort location is always required before ambient-eUV can automatically be considered an instrument.

In summary, this study illustrates how an environmental exposure (ambient-eUV) may be modelled using simple linear regression from routine meteorological data (sunlight). This method provides an attractive alternative to prospective randomized designs, which can be costly and time consuming, and could make it possible to utilize existing data from established cohort studies. In addition it illustrates how the causal link between ambient-eUV, eUV exposure and the production of vitamin D may be able to provide robust insights into the causal effects between vitamin D and a number of different health outcomes via instrumental variable analysis.

Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council, the Wellcome Trust and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors who serve as guarantors for the contents of this paper.

Conflict of interest: None declared.

References

25. Davies M, Mawer EB, Hann JT et al. Seasonal changes in the biochemical indices of vitamin D deficiency in the elderly: a comparison of people in residential homes,


