Historical migration and contemporary health

Andersen, Thomas Barnebeck; Dalgaard, Carl-Johan Lars; Volmar Skovsgaard, Christian; Selaya, Pablo

Published in: Oxford Economic Papers

DOI: 10.1093/oep/gpaa047

Publication date: 2021

Document version: Accepted manuscript


Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying this open access version.

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 26. Oct. 2023
Historical migration and contemporary health

Thomas Barnebeck Andersen\textsuperscript{a}, Carl-Johan Dalgaard\textsuperscript{b}, Christian Volmer Skovsgaard\textsuperscript{c}, and Pablo Selaya\textsuperscript{d}

\textsuperscript{a}Department of Business and Economics, University of Southern Denmark, Campusvej 55, 5230-DK Odense, Denmark, barnebeck@sam.sdu.dk [Corresponding author]

\textsuperscript{b}Department of Economics, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K, Denmark, carl.johan.dalgaard@econ.ku.dk

\textsuperscript{c}Danish Centre for Health Economics, University of Southern Denmark, J.B. Winsløws Vej 9, DK-5000 Odense, Denmark, chsko@health.sdu.dk

\textsuperscript{d}Department of Economics, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K, Denmark, pablo.selaya@econ.ku.dk

Abstract

We argue that migration during the last 500 years induced differences in contemporary health outcomes. The theory behind our analysis builds on three physiological facts. First, vitamin D deficiency is directly associated with higher risk of all-cause mortality. Second, the ability of humans to synthesize vitamin D from sunlight (UV-R) declines with skin pigmentation. Third, skin pigmentation is the result of an evolutionary compromise between higher risk of vitamin D deficiency and lower risk of skin cancer. When individuals from high UV-R regions migrate to low UV-R regions, the risk of vitamin D deficiency rises markedly. We develop a measure that allows us to empirically explore the aggregate health consequences of such migration in a long historical perspective. We find that the potential risk of vitamin D deficiency induced by migration during the last half millennium is a robust predictor of present-day aggregate health indicators.

\textbf{JEL classification:} I1, J1, J15.
1 Introduction

More than 117 million foreign individuals resided in the OECD area in 2013 (OECD, 2015), with the majority of them (70%) born in the South. Total migration into OECD countries increased by almost 40% between 2000 and 2013. Since 2015 European OECD countries have collectively received more permanent migrants than the United States (OECD, 2019). International migration will likely become one of the most important demographic trends of the 21st century, not least in Europe.

In this study we explore the potential long-run health ramifications of international migration. Our starting point is the Age of Discovery, which eventually led to a considerable reshuffling of the world population. We ask whether movements of people between the years 1500 and 2000 hold explanatory power vis-à-vis current aggregate health outcomes across the world at large and across US states. We find that they do.

The theory behind our analysis is related to the risk of premature death and builds on three physiological facts. First, vitamin D deficiency is directly associated with an increased risk of all-cause mortality. Second, the ability of humans to synthesize vitamin D from sunlight (i.e., ultraviolet radiation, UV-R) declines as the intensity of skin pigmentation increases. Third, the intensity of pigmentation is the result of an evolutionary compromise between the costs of pigmentation (e.g., higher risk of vitamin D deficiency) and the benefits (e.g., lower risk of skin cancer), for which reason people from high UV-R regions have more intense skin pigmentation. Accordingly, when individuals indigenous to high UV-R regions (usually the South) relocate to low UV-R regions (usually the North), the risk of vitamin D deficiency rises markedly and with it the potential detrimental impact on average health in the recipient region. In our empirical analysis we find that negative health consequences can be detected in aggregate cross-country data and in independent data across US states.

These findings are of interest from a global health policy perspective. It is well known that the world population is aging, which puts pressure on health expenditures in the years to come. Our estimates suggest that international migration may put additional pressure on expenditures. However, and in stark contrast to aging, the health consequences of South-North migration that we focus on can be mitigated at a relatively modest cost. The cost of a yearly dose of vitamin D (600 IU/day) amounts to about USD 18.1 Consequently, our results suggest that dietary

---

1Holick et al. (2011) recommend at least 400 IU (1 IU = 25 ng) for children 0-1 year; at least 600 IU for children 1-18 years; at least 600 IU for adults 19-50 years; and 600-800 IU for adults 50+ years. At Walgreens
vitamin D supplements offer a low-cost way to reduce both morbidity and premature mortality in populations susceptible to vitamin D deficiency.\(^2\)

In order to analyse the aggregate health implications of historical migration, we construct a measure that proxies the risk of vitamin D deficiency in a given population. The measure tracks the difference between UV-R intensity in the ancestral place of residence of the population (\textit{ancestor} UV-R) and the actual level of UV-R intensity at the place of current residence (\textit{ambient} UV-R). The computation of ancestor UV-R is made possible by using the World Migration Matrix between 1500 and the present (Putterman and Weil, 2010). For each country the database records the fraction of the current population with ancestral (i.e., year 1500 C.E.) origin elsewhere. Ancestor UV-R can therefore be defined as a weighted average of UV-R in ancestral homelands, with 1500 C.E. population shares used as weights. We derive a similar measure for each of the 51 US states. The idea behind our measure is that ancestor UV-R (through a well-established evolutionary mechanism described in Section 2) should capture the average intensity of skin pigmentation in the current population. If ancestor UV-R is high, we expect a population with intense skin pigmentation. Insofar as ancestor UV-R exceeds ambient UV-R in a given location, we expect an elevated risk of vitamin D deficiency in the present-day population, which works to elevate mortality and morbidity, ceteris paribus.

With the difference between ancestor and ambient UV-R as a measure of the potential risk of vitamin D deficiency, we examine its explanatory power vis-à-vis life expectancy across the world.\(^3\) Our baseline finding is that greater risk of vitamin D deficiency is negatively correlated with longevity. This holds true even when we control for well-studied contemporary correlates of longevity such as per capita income, human capital, and income inequality. We also show that neither the level of UV-R in itself nor latitude at the place of residence render the risk-of-vitamin-D-deficiency variable insignificant. This is consistent with identification coming from the imbalance between ancestor and ambient UV-R exposure and not simply from differences in the current environment. Moreover, the (partial) correlation remains significant even when we filter out other historical and geographic correlates of contemporary life expectancy.

\(^2\)Aside from the obvious welfare gains from improved public health, appropriate public health policy (e.g., information campaigns) would help mitigate any induced fiscal sustainability issues.

\(^3\)The relevance of our measure depends on the quality of the migration matrix. It also requires that millennia of human interactions between individuals from different ethnic groups have not ‘muted’ the evolutionary signal from ancestral UV-R on present-day skin complexion too much. Note, however, that if these requirements are not met, the ensuing results will likely be systematically attenuated.
We provide a series of consistency checks of our baseline findings. First, we establish that our risk-of-vitamin-D-deficiency variable turns statistically and economically insignificant when the share of present-day citizens that are natives increases, as we would expect. Second, we re-examine the link between our risk-of-vitamin-D-deficiency variable and longevity within a country with considerable historical immigration: the United States. By relying on within-country variation, the scope for omitted variables bias is much more limited than in the analysis across countries. Conditional on contemporary correlates of longevity, our US cross-state results are qualitative similar to those in the cross-country setting. Third, in the US analysis we are able to provide a more direct measure for actual vitamin D deficiency in each individual state. This allows us to gauge whether our risk-of-vitamin-D-deficiency variable is in fact operating through vitamin D deficiency. The results suggest that this is the case.

Our research is related to the large literature that – following the seminal work of Preston (1975) – studies socioeconomic correlates of longevity. Preston (1975) documented a strong income gradient in longevity across countries but observed that most of the increase in longevity over time was due to upward shifts in the curve (nowadays labelled the the Preston curve), not movements along it. Subsequent research has devoted considerable effort to examining whether indeed income per se is ‘protective’ of mortality. Similarly, a number of contributions have investigated potential determinants of the residual variation in longevity. The list includes education (e.g., Clark and Royer, 2013), inequality (e.g., Deaton, 2003), health technology diffusion (e.g., Soares, 2007; Hansen, 2013), and climate (e.g., Deschenes and Moretti, 2009). Perhaps the most closely related study is Galor and Moav (2007), who hypothesize that the ancestral timing of the Neolithic revolution carries explanatory power in terms of cross-country differences in longevity.\(^4\)

To our knowledge, the present study is the first to document a link between an increased risk of vitamin D deficiency and present-day health differences across countries. Our contribution is twofold: (i) We construct a theoretically meaningful measure of cross-country differences in the potential level of vitamin D deficiency within contemporaneous populations, and (ii) we document a strong link between potential vitamin D deficiency and health differences across and within populations. The study highlights potential health consequences of historical migration flows that hitherto have not been noticed.

\(^4\)The theory is that the Neolithic revolution influenced the nature of the environmental hazards confronted by human populations, thereby unleashing an evolutionary process with observable consequences in terms of cross-country variation in longevity today. See Barnes et al. (2011) for independent support.
Compared to existing evidence within epidemiology that typically derives from less aggregated approaches, we do not offer evidence that directly improves identification of the health impact of vitamin D deficiency. But our study complements the literature in useful ways. In a leading textbook on health disparities in multicultural societies, Bhopal (2014) notes that in the context of vitamin D – but also more generally – the question of causality poses great difficulties because health disparities confound cultural, socioeconomic, and genetic factors. By combining many different socioeconomic and cultural contexts, an aggregate approach can reduce that type of confounding. Moreover, an aggregate approach trades off internal and external validity differently than a microeconometric approach. Consequently, by offering triangulation evidence, an aggregate approach importantly complements more microeconometric approaches to understanding differences in health.

Finally, by exploiting UV-R related variation the present study is related to a study by Andersen et al. (2016) that documents a detrimental impact from the level of UV-R on contemporary prosperity. As shown below, our risk-of-vitamin-D-deficiency measure remains significant when the channel explored in Andersen et al. (2016) is filtered out.

The rest of the paper is structured as follows. Section 2 provides an organizing theoretical framework for the present study. Section 3 provides the cross-country analysis, whereas section 4 revisits the hypothesis by employing data pertaining to US states. Section 4 also presents evidence on the mechanism that we believe drives the main results. Finally, section 5 concludes.

2 Organizing framework

Skin color is a compromise between the evolutionary costs and benefits from UV-R (Jablonski and Chaplin, 2000; Diamond, 2005). On the one hand, UV-R damages the skin in a number of ways. It suppresses sweating and disrupts thermoregulation, as sunburn damages sweat glands. In tropical climes, it increases the risk of infection in sunburned skin. It leads to nutrient photolysis – in particular photolysis of folate – and UV-R may eventually cause skin cancer. Darker skin protects humans against these harmful effects of UV-R. On the other hand,
increased levels of melanin in darker skins increase the risk of vitamin D deficiency. While melanin acts as a sunscreen preventing the harmful aspects of UV-R, as described above, UV-R also converts so-called 7-dehydrocholesterol to pre-vitamin D, which is turned into vitamin D. This process occurs at higher efficiency in less intensely pigmented skin. This is important as exposure to sunlight is the most important source of vitamin D (Pearce and Cheetham, 2010). When the duration of UV-R exposure is insufficient to catalyze pre-vitamin D, individuals are at risk of vitamin D deficiency, which comes at a cost of greater mortality.

Most tissues and cells in the human body have a vitamin D receptor, which possesses the enzymatic apparatus to convert the primary circulating form of vitamin D into an active form. This discovery has led to many new insights into the vital functions of vitamin D (Holick, 2007). Rickets in children and osteomalacia in adults are classic manifestations of profound vitamin D deficiency. However, low vitamin D levels in humans also increase the risk of many non-musculoskeletal conditions such as cancer, metabolic syndrome, diabetes, cardiovascular diseases, as well as infectious and autoimmune disorders (Holick, 2007; Kulie et al., 2009; Pearce and Cheetham, 2010). A meta-study concluded that vitamin D deficiency is directly associated with an increased risk of (all-cause) mortality (Melamed et al., 2008). Hence, there are also costs associated with greater skin pigmentation.

The evolutionary optimal level of skin pigmentation balances costs and benefits of pigmentation (Jablonski and Chaplin, 2000; Diamond, 2005). Formalizing these insights in a simple organizing framework allows us to rationalize our empirical measure of the risk of vitamin D deficiency. Consequently, let $B_s(u)$ capture the benefit in terms of higher life expectancy resulting from the decreased risk of skin cancer associated with darker skin, $s$, for a given level of ambient UV-R, $u$. That is, $B'_s(s, u) > 0$. Likewise, let $C_s(u)$ capture the disadvantage in terms of reduced life expectancy resulting from the increased risk of vitamin D deficiency associated with darker skin, $s$, for a given level of ambient UV-R. That is, $C'_s(s, u) > 0$. The following two additional assumptions are natural: $B'_s(s, u) < 0$, since (for a given level of skin pigmentation, $s$) the risk of skin cancer goes up when $u$ goes up; $C'_s(s, u) < 0$, since (for a given level of skin pigmentation, $s$) the risk of vitamin D deficiency goes down as $u$ increases. Finally, we assume as a melanoma.

---

8Fortunately, excess pre-vitamin D and vitamin D are destroyed by sunlight, for which reason excessive exposure to UV-R does not cause vitamin D intoxication (Holick, 2007).

9A low level of serum 25(OH)D, the main form of circulating vitamin D, is the key marker of vitamin D deficiency (Garland et al., 2006).

10For a broad evidence-based review, see Kulie et al. (2009); for the specific link between vitamin D and cancer, see Garland et al. (2006).
To maximize life expectancy, evolution has determined skin color \( s \) according to

\[
s^* = \arg \max_s LE(s, u),
\]

where \( LE(s, u) \equiv B(s, u) - C(s, u) \), and where \( u \) is treated as a parameter. The (necessary and sufficient) FOC is

\[
B'_s(s^*, u) - C'_s(s^*, u) = 0,
\]

which determines \( s^* \) implicitly as a function of \( u \). By the implicit function theorem we get

\[
\frac{ds^*(u)}{du} = -\frac{B''_{su}(s^*, u) - C''_{su}(s^*, u)}{B''_{ss}(s^*, u) - C''_{ss}(s^*, u)},
\]

which is indeterminate. However, it holds empirically that skin turns darker with UV-R (Jablonowski and Chaplin, 2000; Diamond, 2005), for which reason we impose \( B''_{su}(s^*, u) > C''_{su}(s^*, u) \) to ensure \( \frac{ds^*(u)}{du} > 0 \). In fact, \( B''_{su}(s, u) > 0 \) appears reasonable since the marginal benefit of darker skin (e.g., reduced skin cancer) in terms of life expectancy likely goes up when \( u \) goes up. Moreover, \( C''_{su}(s, u) < 0 \) appears reasonable since the marginal cost of darker skin (e.g., increased vitamin D deficiency) probably goes down as \( u \) increases.

Evolutionary (optimal) life expectancy (the value function) becomes

\[
LE^*(u) \equiv B(s^*(u), u) - C(s^*(u), u).
\]

What is the impact on life expectancy if we compare societies situated in different UV-R environments? That is, what is the impact of changing UV-R slightly from \( u \) to \( u + du \)? The envelope theorem tells us what happens to life expectancy (the value function) when UV-R (the parameter) changes infinitesimally. Since we are initially at an optimum, it is only the direct change that matters; the indirect change working through \( \frac{ds^*(u)}{du} \) is negligible in a neighborhood of the optimum. This means that we are treating \( s^* \) as a constant while infinitesimally varying \( u \). Applying the envelope theorem leads to

\[
\frac{dLE^*(u)}{du} = B'_u(s^*, u) - C'_u(s^*, u),
\]

which is indeterminate. Note, however, that if \( B'_u(s^*, u) = C'_u(s^*, u) \) then evolutionary life ex-
pectancy is independent of UV-R, meaning that in pre-historical times there would be no UV-R induced differences in longevity across (neighboring) traditional societies.

In optimum, there is still a risk of vitamin D deficiency as a result of the trade-off with the harmful effects of UV-R described above. However, these risks are likely to increase appreciably when humans with intensely pigmented skins that evolutionary were adapted to life in high UV-R regions relocate to low UV-R regions. In fact, individuals with highly pigmented skin, who migrate to low UV-R areas, need to increase the exposure time to UV-R up to 10-fold to obtain the same level of vitamin D synthesis as their lighter skinned counterparts (Pearce and Cheetham, 2010). For this reason, such immigrants are at high risk of vitamin D deficiency (Jablonski and Chaplin, 2000; Pearce and Cheetham, 2010). The epidemiological literature has already documented such consequences. Fogelman et al. (1995), for example, have shown that recent migrants from Ethiopia to Israel suffer from vitamin D deficiency and its aforementioned manifestations. Similar results are found among migrants from India to the UK (Henderson et al., 1987). For the US, Ginde et al. (2009) report that (during 2001-2004) 97 percent of all non-Hispanic blacks and 90 percent of all Mexican-Americans had vitamin D deficiency, whereas the comparable share for Caucasians was 70 percent. Garland et al. (2006) cite a number of studies showing that blacks in the US have vitamin D levels half of that of Caucasians.

The formal analysis of the impact on life expectancy of migration to another UV-R environment would be identical to the one conducted above. In particular, we would change $u$ slightly and study the impact on $LE$, keeping $s^*$ constant. The latter is obviously uncontroversial as skin pigmentation is constant in a non-evolutionary time frame. Without a new assumption the comparative static result would therefore be the same as above.

Our new assumption is that in the modern era the evolutionary benefit of intense skin pigmentation is below its pre-historic levels, for three reasons: First, there is widespread awareness of the harmful effects of excessive exposure to sunlight, which leads people to actively try to prevent sunburn.\footnote{It is worth noting that sunscreen with a factor of 15 decreases the synthesis of vitamin D by 99 percent (Ginde et al., 2009).} Second, present-day individuals in low UV-R regions tend to spend much more time indoors than what was the case pre-historically, in effect lowering UV-R exposure.\footnote{The fact that low UV-R regions tend to be richer today, which generally implies fewer jobs in outdoor activities such as agriculture, is documented in Andersen et al. (2016).} Third, effective treatment of skin cancer is now available. As a consequence, the risk of premature death due to excessive sun exposure is likely much below evolutionary levels; in part
because of development itself (fewer outdoor activities), and in part because of active prevention (sunscreen) and the availability of medical treatment of skin cancer. At the same time, it is worth observing that lower exposure times to sunlight, due to changes in the occupational structure in the course of development, and the emergence of effective countermeasures to sunburn both work to increase the risk of vitamin D deficiency. Ceteris paribus, we therefore expect that the cost of intense skin pigmentation is, if anything, above pre-historic levels.

In order to capture the hypothesized shift in relative costs and benefits in a simple way, suppose benefits and costs are a function of the ‘state of development’, $d$. If so, the argument is that while $B_u'(s^*, u; d') - C_u'(s^*, u; d) = 0$ may have held in pre-historical times, today, where $d' > d$, we have $B_u'(s^*, u; d') - C_u'(s^*, u; d') > 0$ (recalling that $C_u'$ and $B_u'$ are both negative). Therefore

$$
\frac{dLE(s^*, u)}{du} = B_u'(s^*, u; d') - C_u'(s^*, u; d') > 0,
$$

and so we conclude that North-South migration ($du > 0$) increases life expectancy, while South-North migration ($du < 0$) does the opposite. More precisely, noting that the comparative static speaks to a change in the environment away from an evolutionary equilibrium, we could proxy $du$ as $du + u^\text{ancestor} - u^\text{ambient}$, where ‘ambient’ refers to UV-R where one resides today and ‘ancestor’ refers to UV-R where one’s ancestors resided. South-North migration has $u^\text{ancestor} > u^\text{ambient}$, which means $du < 0$, whereas North-South migration has $u^\text{ancestor} < u^\text{ambient}$, which means $du > 0$.

Consider South-North migration, which today is the most pertinent type of migration flow. As the difference between the UV-R level of ancestors and the ambient UV-R level expands (and with it the imbalance between actual skin pigmentation and its evolutionary optimal level), the risk of vitamin D deficiency rises above its evolutionary determined level. With North-South migration individuals with ancestors in low UV-R regions move to high UV-R regions, for which reason the risk of vitamin D deficiency declines below the evolutionary determined level associated with local skin pigmentation. In the end, one would expect low UV-R areas with more individuals with ancestry in high UV-R regions to be characterized by lower life expectancy, ceteris paribus, and vice versa in high UV-R regions with a substantial number of individuals with ancestry in low UV-R regions.

The discussion in this section suggests a monotonic link between $(u^\text{ancestor} - u^\text{ambient})$, a difference which we label DIFFUV below, and life expectancy.\textsuperscript{13} DIFFUV > 0 (i.e, $u^\text{ancestor} >$

\textsuperscript{13}We document below that this link is not sensitive to the inclusion of observed mortality risks from skin

9
u_{ambient}^{\text{ambient}}$) captures the discrepancy between actual skin pigmentation and its evolutionary optimal level, which results from South-North migration. Consequently, we expect DIFFUV to be negatively correlated with life expectancy.

3 Cross-country analysis

3.1 Data: Main independent variable

This section explains how we generate DIFFUV by exploiting international movements of people that occurred during the last 500 years. We are interested in constructing a variable that provides us with a measure of a population’s susceptibility to vitamin D deficiency given local climatic conditions. Consequently, we need a measure of local exposure to UV-R. NASA produces daily satellite-based data for UV-R exposure. This UV-R index captures the strength of radiation at a particular location, and it is available in the form of geographic grids and daily rasters with pixel size of 1-degree latitude by 1-degree longitude. The index is the end result of a rather complex calculation. It takes as inputs total ozone column, the earth-sun distance, solar zenith angle, surface irradiance under clear skies, cloud optical thickness, and cloud attenuation factor, which all can be determined at a high resolution.\footnote{NASA expresses the resulting UV-R index in units that speaks directly to how exposed people are to sunburn as a consequence of UV-R. Put differently, the variable becomes an ‘index of the potential for biological damage due to solar irradiation given the local column ozone amount and cloud conditions on each day.’ The measure is explained in more detail in Andersen et al. (2016).}

We rely on data for daily local-noon irradiances for 1990 and 2000 to produce average yearly UV-R levels for each country.\footnote{Though we invoke an average, the correlation between UV-R in 1990 and 2000 is above 0.99. In general, it seems that the intensity of surface UV-R has been relatively stable on earth during the last two billion years (Cockell and Horneck, 2001). Hence in a cross-section context current comparative UV-R levels are likely to be an excellent indicator of UV-R conditions a few centuries ago.} That is, in our analysis below we employ an average of the 1990 and 2000 observations. Figure 1 provides a map depicting the global distribution of UV-R intensity.

cancer.

\textit{Solar zenith angle} (SZA) is the angle between the local zenith (i.e., directly above the point on the ground) and the line of sight from that point to the sun. This means that the higher the sun is in the sky, the lower the SZA is. \textit{Optical thickness} is a measure of the fraction of UV radiation that is not absorbed on a path. Clouds are formed by small water droplets or ice crystals, so UV-R is scattered when passing through them, resulting (in general) in extinction or diminished transmissivity of the atmosphere (Calbó et al., 2005). \textit{Attenuation} depends on different cloud properties in complicated and partly unknown ways.
As explained in section 2, skin color represents an evolutionary compromise between skin types light enough to permit UV-R penetration for vitamin D synthesis but dark enough to reduce harmful effects of UV-R. To permit vitamin D synthesis, this evolutionary compromise results in paler skins at higher latitudes where levels of UV-R are lower. Migration perturbs this evolutionary balance in the sense that humans with darker skins who migrate to high-latitude areas need to increase their exposure time to UV-R up to 10-fold in order to get the same level of vitamin D synthesis as fair skinned individuals. It is crucial that the variable we construct captures this migration-induced perturbation.

Consequently, we ‘ancestry adjust’ NASA’s UV-R variable using the migration-matrix methodology of Putterman and Weil (2010). They construct a matrix detailing the year-1500 origin of the current long-term residents of almost every country in the world. Whenever possible, Putterman and Weil rely on genetic evidence. In cases where direct genetic evidence was not available they used textual accounts and/or generalizations from countries with comparable histories for which genetic evidence were obtainable (e.g., archives on the slave trade, national censuses, and estimates of where the world’s Ashkenazi Jews and Gypsies lived in 1500) in order to map people with such ethnic identifications to particular countries today.

Let $\mathbf{M}$ denote the resulting (square) migration matrix and $\mathbf{m}_i$ the $i$’th row of this matrix, where $\mathbf{m}_i$ is a 1 by $N$ vector. The row vector $\mathbf{m}_i$, which has entries that sum to one, holds infor-
Figure 2: Distribution of DIFFUV across the Globe

Notes: See Andersen et al. (2016) for details on construction of the UV-R index.

mation on the proportion of the ancestors of long-term residents of country $i$ that is estimated to have lived in each source country in 1500. Row 43, for example, is Denmark. The row vector $m_{43}$ has five nonzero entries, corresponding to the five source countries for the current Danish population: Denmark (0.977), Germany (0.005), Netherlands (0.005), Turkey (0.008), and Yugoslavia (0.005).

Our independent variable of interest is

$$\text{DIFFUV}_i \equiv m_i \text{UV} - \text{UV}_i,$$  \hspace{1cm} (1)

where $\text{UV}$ is the $N$ by 1 vector of UV-R levels and $\text{UV}_i$ is UV-R in country $i$. DIFFUV$_i$ then measures the difference between ancestry adjusted UV-R in country $i$ and the actual level of UV-R in the said country. This means that a country $i$ with DIFFUV$_i > 0$ has a population that (on average) has a skin type more suited for living in places with stronger UV-R (as measured by $m_i \text{UV}$) than in the country $i$ where they actually live (as measured by $\text{UV}_i$). That is, DIFFUV$_i > 0$ means that the population living in country $i$ has difficulties getting sufficient vitamin D since their skin pigmentation is evolutionary adapted to a place with higher UV-R.

Figure 2 provides a map of the DIFFUV variable. Eyeballing the figure immediate reveals that European offshoots (e.g., Australia and countries in South America) have large negative
values of DIFFUV, as expected. In fact, Australia (-83.1) is second from the bottom in the DIFFUV distribution. Only Singapore (-89.3) scores lower than Australia. The explanation is that Singaporeans are 77 percent ethnic Chinese; and with UV-R in China being only 56.4 percent of the level of UV-R in Singapore, the outcome is given. On the other hand, Eurasian countries in general have positive values of DIFFUV. Most African countries as well as the Indian subcontinent have values of DIFFUV close to zero; the southernmost region of Africa, however, tends to have DIFFUV < 0. The highest level of DIFFUV is also African: viz. Swaziland (37.1). The explanation is that 94.15 percent of the population of Swaziland is ethnic Mozambican, and UV-R in Swaziland is only 82.6 percent of the level of UV-R in Mozambique.

3.2 Empirical strategy

The specification we take to the cross-country data is the following:

$$H_i = \beta_0 + \beta_1 \cdot \text{DIFFUV}_i + Z_i' \cdot \gamma + \epsilon_i,$$

where $H_i$ is the health indicator in country $i$ in the year 2010; DIFFUV$_i$ is the difference between ancestry adjusted UV-R and actual UV-R in country $i$, as defined by equation (1); $Z_i'$ is a vector of auxiliary controls. Our primary health indicator in the analysis below is life expectancy at birth.\(^\text{16}\) However, we also examine the link between DIFFUV and adult mortality, mortality for children under the age of five, and infant mortality. In all settings the parameter of interest in equation (2) is $\beta_1$.\(^\text{17}\)

When estimating equation (2) by OLS our main concern is omitted variables, as many different factors may influence life expectancy. In an effort to minimize the risk that our results are driven by omitted factors, we employ a relatively extensive baseline specification.

Our baseline controls involve three key determinants of health and a full set of (eight) regional fixed effects. Identification of an influence from DIFFUV is thus sought in the within

\(^{16}\)We use the log of life expectancy, but unlogged life expectancy produces similar results, cf. tables B5-B8 of the online appendix. The correlation between logged and unlogged life expectancy in the global sample is more than 0.99.

\(^{17}\)It may be worth emphasizing that we are essentially pursuing a proxy variable approach, where DIFFUV proxies the average susceptibility in the present-day population to vitamin D deficiency. This requires that the ancestral UV-R measure is a meaningful marker of average skin pigmentation in present-day populations. If the migration matrix is of poor quality, and/or if interbreeding over time has dulled the underlying evolutionary signal, ancestral UV-R will not be a meaningful marker. In this case, we normally expect an attenuation bias that goes against rejecting the null hypothesis, $H_0 : \beta_1 = 0$. 

13
regional variation in health outcomes. Following Preston (1975) and many subsequent contributions we include GDP per capita. Naturally, there are good reasons why income could be a co-determinant of health via nutrition, health investments, or both. In the present case, however, there is an additional reason why it seems worth controlling for average income. As economic development likely reduces the exposure time to UV-R via secular changes in the occupational structure, the inclusion of GDP per capita sharpens the interpretation of DIFFUV. We thereby ask whether DIFFUV can account for some of the residual variation in health around the Preston curve, as it should in theory.

The baseline Preston curve is further augmented, however, in that our baseline set of controls (in addition to regional fixed effects) also involve income inequality (cf., Deaton, 2003; Ram, 2006) and human capital in the form of schooling (cf., Grossman, 2006). Naturally, all these variables are quite likely endogenous vis-à-vis life expectancy, which we try to address by lagging all variables by roughly a decade; life expectancy is measured in 2010, whereas our controls are measured in 2000 at the latest. These controls have the virtue that they collectively account for a large share of the variation in global health, for which reason we indirectly control for a variety of mechanisms. In this manner we aim to reduce the risk that our main results are tainted by omitted variables bias.

In further tests we explore the robustness of the DIFFUV indicator to the inclusion of additional (potential) confounders, as discussed below. Summary statistics of the dependent variable and the confounders employed in the analysis, as well as their sources, are provided in the online appendix, section A.

3.3 Empirical results

3.3.1 Baseline findings

Table 1 reports our baseline results. Observe that all estimates are standardized, which means that individual point estimates provide us with the number of standard deviations that the left-hand-side variable changes when the associated right-hand-side variable increases by one stan-

---

18The regions are: East Asia and the Pacific; Eastern Europe and Central Asia; Middle East and North Africa; Western Europe; South Asia; Sub-Saharan Africa; North America, and Latin America.
19See Baird et al. (2011) for evidence of the income-mortality link in developing course; Deaton (2003) for an overview; and Dalgaard and Strulik (2014) for theoretical foundations for the income-longevity nexus and the concavity of the Preston curve.
20Specifically, income inequality is measured as an average from 1960-1996. The remaining variables are measured in 2000.
standard deviation. Moreover, p-values (reported in parenthesis) are based on robust standard errors.

TABLE 1

In the first six columns we estimate unrestricted versions of the baseline specification, where ancestor and ambient UV-R enter separately in the model. As one moves from left to right, we progressively add controls; column 6 include per capita income, inequality, schooling, and regional fixed effects simultaneously.

As expected, ancestor and ambient UV-R enter with opposite sign; conditional on ambient UV-R, higher ancestor UV-R is associated with lower life expectancy. The final row of the first six columns reports the p-values associated with the ‘structural test’ that the two point estimates are identical in absolute value. Inspection of the table reveals that this hypothesis cannot be rejected as soon as we add to the model (in addition to regional fixed effects) respectively per capita income, inequality, schooling, or all three variables simultaneously.

The next six columns estimate the restricted model, which involves DIFFUV; i.e., equation (2). As expected, the baseline model accounts for the lion’s share of the variation in longevity, with an adjusted R-squared between 0.79 and 0.84 depending on the exact specification. As can be seen, our measure of risk-of-vitamin-D deficiency is statistically significant in all settings. Economically, the effect of DIFFUV is substantial in that the standardized point estimates are of the same order of magnitude as income, inequality, and schooling, respectively. At the same time, however, the variation in DIFFUV is considerably smaller than the corresponding variation in the other three determinants, for which reason the fraction of the total variation in longevity that can be accounted for by DIFFUV is modest in comparison. In the simple univariate case (column 7), DIFFUV only accounts for about five percent of the total variation in life expectancy.

3.3.2 Auxiliary controls

The baseline model obviously accounts well for the variation in longevity. However, one may wonder if other well-known determinants of health can account for the part of the residual variation in the augmented Preston curve that is now accounted for by DIFFUV.

TABLE 2
In table 2 we examine a further set of contemporary determinants of health. First and foremost, the table demonstrates that mortality due to skin cancer does not influence the point estimate for DIFFUV appreciably. Consistent with priors, skin cancer is not significant conditional on DIFFUV and our baseline set of controls. The remaining columns examine whether adding health expenditure, alcohol consumption, air pollution, the urbanization rate, a measure of gender equality, or political institutions influence the link between health and DIFFUV. In column 8 we include all controls at once; and in column 9 we drop the baseline controls, which increases the sample size considerably. The basic message from table 1 carries over to table 2.

Table 3 shifts attention to geographic and historical determinants of health and examines their influence on the DIFFUV/longevity link in a similar fashion to that of table 2.

**TABLE 3**

As expected from the tests in table 1, table 3 (column 1) shows that ambient UV-R does not influence the DIFFUV estimate, conditional on our baseline controls. This demonstrates that the effect we are presently capturing is due to the imbalance between local and ancestor UV-R; and not just the former, which might impact life expectancy via per capita income (Andersen et al., 2016).

Column 2 addresses another concern. UV-R is strongly correlated with the distance to the equator, which in turn is correlated with economic development. Hence, positive (negative) values of DIFFUV implies immigration from relatively poor (rich) countries. This raises the question of whether an influx of immigrants from poor (rich) places is what accounts for our result that positive (negative) levels of DIFFUV lower (increases) longevity in the host country, rather than the proposed mechanism involving skin pigmentation? If prosperity influences the amount of ‘health capital’ an individual is endowed with, this alternative link is difficult to rule out a priori. Thus, in an effort to gauge whether this alternative channel is operative, column 2 controls for DIFFABSLAT, which is the difference between the ancestral absolute latitude of the current population and the absolute latitude of the individual country. When both DIFFUV and DIFFABSLAT are included the former variable captures differences in UV-R as explained by elevation and cloud cover, the two key determinants of UV-R aside from absolute latitude.

---

21These controls are pretty standard in the literature on the ‘correlates of health’; see, e.g., Or (2000), Moore and Keary (2003), and Marmot (2006). See, however, Besley and Kudamatsu (2006) on a possible democracy link.

22Naturally, the model already controls for per capita income. However, the latter is very likely measured with error, so UV-R could still emerge significant in the model.
(see section 3.1). Reassuringly, DIFFUV remains significant when DIFFABSLAT is added to the model (see columns 2, 12, and 13).

Overall, the DIFFUV estimate appears fairly robust, which suggests that we should not be concerned about omitted variables.\(^{23}\) At the same time, the table provides some additional support for prior findings in the literature. Specifically, the influence from the ancestor-adjusted timing of the Neolithic revolution (Galor and Moav, 2007) and the genetic proximity to the US (Hansen, 2013) receive support. The remaining controls do not appear to be significant, conditional on our baseline controls. Examining column 13, however, where we omit our baseline controls, we observe that a set of additional controls turn significant. For example, a later timing of the fertility transition, consistent with the theory of Cervellati and Sunde (2005), is associated with lower life expectancy. Similarly, ambient UV-R is significant, consistent with an influence on living standards as proposed by Andersen et al. (2016). With respect to DIFFUV, however, the basic message from table 1 carries over.

### 3.3.3 Other mortality measures

Another interesting issue is how our baseline results unbundle. Do they reflect a link to adult mortality, or perhaps child mortality and/or infant mortality? A priori, we expect that DIFFUV influences health among the young as well as among the old. Medical research has established that vitamin D influences a whole range of diseases, which may afflict young and old. More fundamentally, vitamin D even influences the human immune system (Pearce and Cheetham, 2010).

Turning to the results, table 4 explores the link between DIFFUV and adult mortality, child

---

\(^{23}\)Coefficient stability is not sufficient to evaluate the bias from omitted variables. Coefficient movements need to be scaled by movements of R-squared (Oster, 2019). Oster proposes an estimator that corrects for the bias from omitted variables by performing such scaling. It works as follows: Let \(\hat{\beta}\) be the coefficient on DIFFUV from a regression of life expectancy on DIFFUV and all observed confounders (i.e., column 12 of table 3), and let \(\hat{R}\) be the associated R-squared. Likewise, let \(\hat{\beta}^{*}\) be the coefficient on DIFFUV from a regression of life expectancy on DIFFUV and no confounders, and let \(\hat{R}^{*}\) be the associated R-squared. Finally, let \(R^{\text{max}}\) be the R-squared from a regression of life expectancy on DIFFUV and all controls, observed as well as unobserved. Assuming equal selection (i.e., observables and unobservables have equal impact on DIFFUV), we have that

\[
\beta^{*} = \hat{\beta} - \left[\hat{\beta} - \hat{\beta}^{*}\right] \times \left[\frac{R^{\text{max}} - \hat{R}}{\hat{R} - \hat{R}^{*}}\right],
\]

where \(\beta^{*}\) is the estimator adjusted for omitted variables bias. The specification in column 12 of table 3 has \((\hat{\beta}, \hat{R}) = (-0.34, 0.90)\) (We are using R-squared, not the adjusted version.); and from a regression of life expectancy on DIFFUV, we get \((\hat{\beta}, \hat{R}) = (-0.19, 0.04)\). Inserting these values and assuming that \(R^{\text{max}} = 1\), we obtain \(\beta^{*} = -0.36\). This is very close to the estimate from the long regression, \(\hat{\beta}\). We take it as a further sign that omitted variables are unlikely to be a problem.

---
(<5 years) mortality, and infant mortality, respectively.

**TABLE 4**

In the first three columns we explore a stripped down specification, where we only condition on regional fixed effects; the next three columns involve our full baseline specification; and, finally, columns 7-9 estimate the unrestricted baseline model. The general theme is that DIFFUV correlates in the expected way with all three mortality rates. Statistically, the link is stronger for child mortality and infant mortality than for adult mortality, which is a bit too imprecisely estimated to be deemed statistically significant at conventional levels.\(^\text{24}\) Economically, the point estimates are again of same order of magnitude as each of the baseline controls; slightly larger in fact than the link involving income inequality, but smaller than the those associated with income and human capital. Nevertheless, the estimates quite clearly suggest a link between the risk of vitamin D deficiency and poorer health, even among younger age groups.

### 3.3.4 Interaction effect

The fundamental hypothesis under scrutiny is that historical population movements have produced a mismatch between evolutionary optimal and actual skin pigmentation. By implication, we expect to see a stronger economic impact from DIFFUV in regions that have experienced more immigration. If this is not the case, our findings could reflect some omitted factor that is not accounted for in the checks conducted above.

One way to perform this consistency check of our hypothesis is to introduce an interaction effect between DIFFUV and the share of the population that can be considered native. The partial effect of DIFFUV will then depend on the share of natives. If we find that the partial effect of DIFFUV is diminished as the share of natives rises, this will support our hypothesis. We follow the approach of Wooldridge (2013, pp. 190-191), which entails introducing the interaction by means of a simple reparameterization. By introducing the interaction as \((\text{DIFFUV} - \mu_D^j) \times (\text{NATIVE} - \mu_N^j)\), where \(\mu_D^j\) and \(\mu_N^j\) are the \(j\)’th percentile of each distribution, we obtain that the coefficient on DIFFUV is the partial effect evaluated at the said percentile.\(^\text{25}\) We will consider the 5th (high immigration), 25th, 50th, 75th, and 95th percentiles.

\(^{24}\)What drives this (marginal) insignificance is evident from the final entry in column 7, where we see that the restricted model of column 4 is firmly rejected. Put differently, we cannot impose the restriction that the point estimate of UV-R ancestor equals (in absolute value) the point estimate of UV-R ambient.

\(^{25}\)We also immediately obtain the correct standard errors at the \(j\)’th percentile.
Table 5 reports the results. (Note that each regression includes the set of baseline controls plus regional fixed effects.)

**TABLE 5**

Consider the first row of table 5. It gives the partial effect of DIFFUV on life expectancy at different percentiles of the share of natives distribution. Evaluated at the fifth percentile – the row 1, column 1 entry – the partial effect is negative and statistically significant. At the 25th percentile – the row 1, column 2 entry – the partial effect is numerically smaller but still statistically significant. As we successively consider higher percentiles, the partial effect of DIFFUV becomes (numerically) smaller and turns statistically insignificant. The other mortality measures display similar patterns.

Overall, the results in table 5 are consistent with our hypothesis: when the immigration share falls, the impact from DIFFUV falls. These results suggest that we obtain identification through the reshuffling of world population after 1500 C.E.

### 4 Cross-state analysis

In this section we redo the cross-country analysis using US cross-state data. We examine the link between DIFFUV and life expectancy using the same empirical strategy as in the cross-country analysis.

The cross-state analysis amounts to a further robustness test based on an independent dataset. If the cross-country nexus between DIFFUV and life expectancy (reported above) is also present in independent cross-state data, it strengthens the empirical case in favor of an impact of international migration on contemporary global health differences.

More specifically, analyzing the nexus between DIFFUV and life expectancy within US states is attractive for two main reasons. First, by studying the nexus within a particular country we eliminate any omitted variable bias stemming from country-specific factors. Second, in the US setting we have access to a proxy for vitamin D deficiency, which enables us to check whether our DIFFUV variable is actually capturing the sort of variation that we claim it is.
4.1 Data: Main independent variable

In order to generate DIFFUV across US states we need to construct a US state level migration matrix and calculate average UV-R levels for each state. To construct the migration matrix we turn to the same data used by Putterman and Weil (2010) and follow their methodology.26 The details are laid out in section B of the online appendix. A visual representation of DIFFUV for the US states is provided in figure 3.

Hawaii (-97.91) and Florida (-75.65) stand out with the largest negative values of DIFFUV. Hawaii and Florida also have the highest levels of UV-R radiation. At the other end of the spectrum are District of Columbia (52.21) and Alaska (44.76) with the two highest positive levels of DIFFUV and respectively the median and the lowest level of UV-R. The fact that DIFFUV is so high in District of Columbia, given a median value of UV-R, derives from the fact that almost half of the city’s population is African American.

As is evident from figure 3, the South does not stand out in terms of DIFFUV. However, states like Georgia, Louisiana, and Mississippi do stand out in terms of ancestor UV-R, as is documented in figure 1 in section B of the online appendix. The reason these states do not stand out in terms of DIFFUV is that ambient UV-R is also high.

The details on data sources as well as summary statistics are relegated to section A of the online appendix.

---

26Their main appendix 1.1 describes how they calculate ancestry shares.
4.2 Empirical results

4.2.1 Baseline findings

Table 6 reports the US state level results that correspond to the results reported in table 1 for the cross-country sample. Two differences are worth highlighting. First, in the US sample we employ a full set of nine census regional fixed effects, implying of course that we rely on within census-region variation.\(^{27}\) Second, as our measure of human capital we employ the fraction of the population with at least a bachelor’s degree, rather than average years of schooling. Once again, all estimates are standardized and robust p-values are reported in parenthesis.

TABLE 6

In the first six columns we estimate the unrestricted model, where ambient UV-R and ancestor UV-R enter separately. Much like in the cross-country context, once we condition on baseline controls the data support the restricted version of the model in which we impose that the estimates for ambient UV-R and ancestor UV-R sum to zero.\(^{28}\) The last six columns therefore estimate the restricted model, where the independent variable is DIFFUV. Inspection of the said columns reveals that DIFFUV is significant at the five percent level or better once regional fixed effects are added. The economic significance of DIFFUV is of the same order of magnitude as income, inequality, and human capital; and once again it is worth observing that the variation in DIFFUV is small compared to income, for which reason the contribution to the overall adjusted R-squared is modest. Nevertheless, the results corroborate the cross-country findings. In fact, the (standardized) point estimate in the US setting is somewhat larger than the comparable cross-country estimate, though in the same ballpark.

4.2.2 Auxiliary controls

As in the cross-country setting we have performed a set of checks of the baseline results by examining the resilience of the DIFFUV/life expectancy link to the inclusion of additional

\(^{27}\)The nine census divisions are: New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific.

\(^{28}\)We observe from column 1 that ambient UV-R is insignificant when census fixed effects are excluded. A likely explanation is that life expectancy is lower in the poorer Southern United States, where UV-R exposure is also greater. This introduces an omitted variable bias that may render ambient UV-R insignificant. Moreover, a negative link between life expectancy and ancestral UV-R (which is also relatively higher in the Southern United States) may arise. As a simple check of this explanation, we ran a regression in which we added income to the two UV-R variables (as we do in column 3), without controlling for census fixed effects. In this case ambient UV-R turns negative and significant, as expected.
controls (not reported). In particular, we confirm the results from the cross-country setting that DIFFUV remains significant when we add either ambient UV-R (suggesting that identification is attained via the mismatch between ancestor and ambient UV-R) or skin cancer incidence (supporting a modest impact from this cause of death on aggregate life expectancy today) to the model on top of our baseline controls. We have also examined the consequences of including a DIFFABSLAT variable to the model, as in the cross-country context. While the economic significance of DIFFUV is unaffected, the parameter of interest is estimated with slightly less precision (p-value is 0.107), presumably due to collinearity.  

4.2.3 Vitamin D deficiency and DIFFUV

In the preceding analysis we have not brought evidence to bear on whether DIFFUV is operating through the proposed physiological mechanism, which may give rise to the following concern: perhaps the results are not linked to vitamin D deficiency, but health related consequences of stress induced by status concerns?

The argument is the following. Movements of people since 1500 brought intensely pigmented people to low UV-R regions, where they tended to enter at the bottom of the social hierarchy. In contrast, less intensely pigmented individuals often ended up moving to high UV-R regions where they entered at the top of the social hierarchy as settlers from colonial powers. If the low social standing of people with ancestry in high UV-R regions caused them to be discriminated against, this could unleash adverse health effects (e.g., Williams, 1999) and thus potentially motivate a negative link between DIFFUV and mortality.

We believe that the preceding analysis has already importantly mitigated this concern, as we controlled for DIFFABSLAT and a range of socioeconomic characteristics of relevance to this particular channel: income inequality; ethnic fractionalization; and even alcohol consumption, which can be induced by low social standing and holds adverse health effects (see also section 3.3.2). Nevertheless, doubts may linger.

---

29 Indeed, the variance inflation factor associated with DIFFUV is 11.48. Moreover, if we exclude income per capita, which is insignificant, DIFFUV turns significant with a p-value of 0.099.

30 In the opposite case (i.e., when immigrants from low UV-R regions enter high UV-R regions) it is perhaps not clear what the end result would be in terms of health. The low pigmented settlers would likely experience higher social status than in their homeland, but their presence would almost surely affect stress levels (and status concerns) in the colonized native population. As a result, it is not obvious that the ‘status channel’ should give rise to a monotonic link between DIFFUV and health outcomes.

31 In table B4 of the online appendix, we estimate specifications with regression splines. DIFFUV runs from approximately -90 to 40, and we have therefore inserted knots at -45 and 0. Inspection of table B4 reveals that it is only the positive interval that holds explanatory power for the different mortality measures. This is consistent
There is unfortunately no way of providing direct evidence of vitamin D deficiency at the country level, and at the US state level no direct data on average state-specific, or region-specific, vitamin D status of the population seems to be available (see Camargo et al., 2007). Yet indirect evidence can be brought to bear.

Specifically, in this section we utilize the link between vitamin D deficiency and anaphylaxis; the latter being a serious allergic reaction (often caused by food), which is rapid in onset and may even cause death. A growing body of evidence suggests that vitamin D deficiency is an important cause of anaphylaxis (Mullins and Camargo, 2012). Laboratory evidence, for instance, suggests several mechanisms through which vitamin D affects allergic reactions in general and anaphylaxis in particular (Camargo et al., 2007). Studies also show a clear relationship between season of birth (fall and winter, the least sunny months) and food allergy prevalence (Sharief et al., 2011). A large US survey shows higher rates of food sensitization in infants born to mothers with low vitamin D intake during pregnancy (Nwaru et al., 2010). Finally, several studies document that epinephrine (a medicine used for life-threatening allergic reactions) autoinjector prescription rates vary with latitude (proxy for exposure to sunlight) in Australia, the UK, and the US (see Peroni and Boner, 2013).

Accordingly, we propose to employ epinephrine autoinjector prescription rates (EAPR) as a crude proxy for actual vitamin D deficiency across US states. The questions we are then able to pose are the following: Does DIFFUV predict EAPR? Does EAPR correlate with life expectancy once we omit DIFFUV? Naturally, if both answers are in the affirmative then this further supports the interpretation of our main findings.

**TABLE 7**

Table 7 provides answers to these questions. In the first five columns we explore whether DIFFUV is a predictor of EAPR. In interpreting EAPR as a proxy for health we also control for our baseline variables: income, inequality, and human capital, as well as regional fixed effects. As can be seen upon inspection of the said columns, DIFFUV indeed correlates with EAPR in the expected way.

In the two remaining columns we explore the link between EAPR and life expectancy. Column 6 demonstrates that EAPR is negatively correlated with life expectancy, consistent with the notion that it constitutes a proxy for vitamin D deficiency. Column 7 documents that

---

with the vitamin D channel.
when we add DIFFUV, EAPR turns insignificant. Naturally, if we had actual data on vitamin D deficiency, we would expect this variable to dominate DIFFUV, being a more proximate cause of health problems. In the present case this is not the outcome, perhaps indicating that EAPR is a somewhat noisier proxy for vitamin D deficiency than DIFFUV. Nevertheless, column 7 suggests that EAPR indeed captures the same sort of variation that DIFFUV is picking up, which likely is vitamin D deficiency. We thus interpret these results as supporting that DIFFUV is picking up variation in the risk of vitamin D deficiency, as befits our hypothesis.

5 Conclusion

We have examined whether a migration-induced imbalance between the intensity of skin pigmentation and ambient UV-R holds explanatory power vis-à-vis present-day global health differences. We find that it does. Consequently, our results suggest that low UV-R regions that have received substantial immigration from high UV-R regions experience lower life expectancy than would have been the case in the absence of such migration flows.

The underlying theory derives from the life sciences. Conditional on ambient UV-R, individuals with intense skin pigmentation (deriving from high ancestral UV-R exposure) are more susceptible to vitamin D deficiency, which is a leading cause of a range of afflictions that cause premature death. The contribution of the present study lies in exploring whether this theory holds explanatory power in the aggregate. The weight of the evidence presented above suggests it does.

While the economic significance of our measure of the risk of vitamin D deficiency (if taken at face value) is relatively strong, it is also clear that its ability to account for cross-country variation in life expectancy is modest. However, if current movements of people continue, which to a large extent represent movements from ‘South to North’, much more variation is likely to become visible during the 21st century. As such, vitamin D deficiency may become an increasing public health issue in the years to come, at least in the absence of preventive public health measures.

We believe the present study could be usefully extended in the direction of studying within country migration. For example, Black et al. (2015) find that the Great Migration within the US reduced the health of African Americans significantly. While the authors suggest that some of the impact may be linked to changes in the intake of alcohol and cigarette smoking, it is worth
noting that migrants also experienced changes in the environment. For example, moving from Georgia to New York would imply a reduction in ambient UV-R of roughly 43%, implying in turn a considerable increase in the risk of vitamin D deficiency for an African American. Whether a vitamin D mechanism could be contributing to the decline in health outcomes in the aftermath of the Great Migration seems to be an interesting topic for future research.

**Funding and conflicts of interest**

We have not received any funding for this project, and we have no conflicts of interest that we are aware of.

**Acknowledgment**

We thank seminar participants at the University of Copenhagen, University of Groningen, University of Southern Denmark, Hitotsubashi University, University of Oxford, and the University of Göttingen for comments.

**Supplementary material**

Supplementary material is available at the OUP webpage. These are the data, the replication do-files, and the online appendix.
References


Williams, D R (1999) “Race, socioeconomic status, and health the added effects of racism and discrimination.”