Historical Migration and Contemporary Health*

Thomas Barnebeck Andersen
Carl-Johan Dalgaard
Pablo Selaya
Christian Volmar Skovsgaard†

August 15, 2018

Abstract

We show that migration during the last 500 years induced differences in contemporary health outcomes. The theory behind our analysis is related to the risks of premature death, and 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 (i.e., ultraviolet radiation, UV-R) declines with skin pigmentation. Third, skin pigmentation is the result of an evolutionary compromise between its costs and benefits (e.g., higher risk of vitamin D deficiency and lower risk of skin cancer), which explains why natives of high UV-R regions became more intensely pigmented. In accord with these physiological premises, when *individuals* indigenous to high UV-R regions migrate to low UV-R regions, the risk of vitamin D deficiency rises markedly and with it the risk of premature death. We develop a measure that allows us to empirically explore the *aggregate* health consequences of such type of migration in a long historical perspective. Our results show 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, both across countries and across US states.

Keywords: Health, longevity, vitamin D, ultraviolet radiation, skin pigmentation, historical migration.

JEL classification: I1, J1, J15.

*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.

†Contact details and institutional affiliations: Andersen (barnebeck@sam.sdu.dk) and Skovsgaard (chsko@sam.sdu.dk) are affiliated with the Department of Business and Economics at the University of Southern Denmark. Dalgaard (carl.johan.dalgaard@econ.ku.dk) is affiliated with the Department of Economics at the University of Copenhagen, CAGE and CEPR. Selaya (pablo.selaya@econ.ku.dk) is affiliated with the Department of Economics at the University of Copenhagen.

1 Introduction

More than 117 million foreign individuals were residing in the OECD area in 2013 (OECD 2015), with the majority of them (70%) born in the South (UN 2014). Total migration into the OECD countries increased by almost 40% between 2000 and 2013, and preliminary data suggest that it accelerated again in 2014 and 2015 (OECD 2015). On these accounts, as well as more recent developments such as the increased number of asylum seekers in Europe and the US (UNHCR 2016), it seems likely that international migration will become one of the most important demographic trends of the 21st century.

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. Concretely, 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 the answer to be in the affirmative.

The theory behind our analysis is related to the risks 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 human skin pigmentation is the result of an evolutionary compromise between the costs of skin pigmentation (e.g., higher risk of vitamin D deficiency) and its 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 such health consequences can be detected in aggregate cross-country data as well as 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 and that this may increase health expenditures in the years to come. Our estimates suggest that international migration may put additional pressure on public (health) finances. However, in stark contrast to aging, the health consequences of South-North migration, due to the mechanism highlighted in the present study, 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,

[1]The Endocrine Society (2011) recommends 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 Wal-

our results suggest that dietary vitamin D supplements offer a low-cost way to reduce both morbidity and premature mortality in populations susceptible to vitamin D deficiency. 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.

In order to analyze the aggregate health implications of historical migration flows, 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 (ancestor UV-R) and the actual level of UV-R intensity at the place of current residence (ambient UV-R). The computation of ancestor UV-R is made possible by using the World Migration Matrix between 1500 and the present, built by Putterman and Weil (2010). For each country the database records the fraction of the current population with ancestral (i.e., year 1500 C.E.) origins 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, one would expect a population with relatively intense skin pigmentation. Insofar as *ancestor* UV-R exceeds *ambient* UV-R in a given location, one would expect an elevated risk of vitamin D deficiency in the present-day population, which should work to elevate mortality and morbidity, ceteris paribus.[2]

With the difference between ancestor and actual UV-R as a measure for the potential risk of vitamin D deficiency, we proceed to examine its explanatory power vis-à-vis life expectancy across the world. Our baseline finding shows that greater risk of vitamin D deficiency is negatively correlated with longevity. This holds true even when we condition on well-studied contemporary correlates of longevity such as per capita income, human capital and income inequality. We also document 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, consistent with the hypothesis that we obtain identification through the imbalance between ancestor and ambient UV-R exposure, not simply from differences in the current environment. Moreover, the correlation we uncover remains significant even when we filter out other historical and geographic correlates of contemporary life

greens (the largest drug retailing company in the US), 100 vitamin D tablets (1 tablet = 1000 IU) cost USD 9.99 (www.walgreen.com was visited 1 December 2015). As the diet will provide some vitamin D, one tablet every second day should suffice, which amounts to about USD 18 for a full year.

[2]One may wonder if other adverse health consequences would arise if the ancestral UV-R falls short of actual UV-R, implying a non-monotonic link between our measure and longevity. We argue in Section 2 that this is unlikely to be the case today, for a variety of reasons.

expectancy.

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 we limit the sample to countries where the bulk of present-day citizens are native, as one would expect. Second, we re-examine the link between our risk-of-vitamin-D-deficiency measure 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, which allows us to gauge whether our risk-of-vitamin-D-deficiency measure is in fact operating through vitamin D deficiency. The results suggest that this is the case.

The present study 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"), and 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 (Soares 2007; Hansen 2013), climate (e.g., Deschênes and Moretti 2009), and more. 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.[3]

To our knowledge, the present study is the first to document a link between an elevated risk of vitamin D deficiency and present-day health differences across countries. Accordingly, the main value added we offer to the literature is twofold: (i) We contruct a theoretically meaningful mesure 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 comparative cross-country health differences. The present study thus highlights potential health consequences of historical migration flows that hitherto have not been noticed.

[3]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.

The present paper does not offer evidence that directly improves upon identification of the health impact from vitamin D deficiency, compared with existing evidence within epidemiology that typically derive from more disaggregated approaches. But the present 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 difficulty 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 recent study by Andersen et al. (2016), which 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 a theoretical organizing 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 today understood as 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[4], and in particular photolysis of folate[5]; and it may eventually cause skin cancer.[6] Darker skin protects humans against these harmful effects of UV-R. On the other hand, increased

[4]Photolysis is the chemical process by which molecules are broken down into smaller units through the absorption of light.

[5]Folate (a vitamin B), or folic acid, is necessary for the formation of normal red blood cells and the synthesis of DNA. Folate is also necessary for normal development of a fetus's nervous system.

[6]Malignant melanoma is by far the most dangerous type of skin cancer, but it is also the least common. There are two other types of skin cancer: basal cell cancer and squamous cell cancer. Basal cell cancer, the most common type of skin cancer, almost never spreads; squamous cell cancer is more dangerous, but not nearly as dangerous as a melanoma.

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-dehydrocholestorol 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).[7] 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.

More concretely, 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).[8] 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). Indeed, a meta-study concluded that vitamin D deficiency is directly associated with an increased risk of (all-cause) mortality (Melamed et al. 2008).[9] 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: $B'_u(s,u) < 0$, since (for a given level of skin pigmentation, s) the risk of skin cancer goes up when u goes up; $C'_u(s,u) < 0$, since (for a given level of skin pigmentation, s) the risk of vitamin D deficiency goes down as u increases. Finally, we assume $C''_{ss}(s,u) > B''_{ss}(s,u)$ and $C'_s(0,u) < B'_s(0,u)$ to ensure an interior optimum.

[7] Fortunately, excess pre-vitamin D or vitamin D is destroyed by sunlight, for which reason excessive exposure to UV-R does not cause vitamin D intoxication (Holick 2007).

[8] A 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).

[9] For 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 (Jablonski and Chaplin 2000; Diamond 2005), for which reason we impose $B''_{su}(s^*,u) > C''_{su}(s^*,u)$ to ensure $\frac{ds^*(u)}{du} > 0$.[10]

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 expectancy 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 apprecia-

[10] In fact, $B''_{su}(s,u) > 0$ appears reasonable since the marginal benefit of darker skin 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 probably goes down as u increases.

bly 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 benefits of intense skin pigmentation is below its pre-historic levels, for three reasons: First, there is acute awareness of the harmful effects of excessive exposure to sunlight, which leads people to actively try to prevent sunburn.[11] 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.[12] 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, one would therefore expect that the cost of intense skin

[11] It is worth noting that sunscreen with a factor of 15 decreases the synthesis of vitamin D by 99 percent (Ginde et al. 2009).

[12] The fact that low UV-R regions tend to be richer today, which generally implies fewer jobs in outdoor activites such as agriculture, is documented in Andersen et al. (2016).

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

$$\frac{dLE(s^*,u)}{du} = B'_u\left(s^*,u;d'\right) - C'_u\left(s^*,u;d'\right) > 0,$$

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, one could proxy du as $du + u^{ancestor} = u^{ambient}$, 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 $u^{ancestor} > u^{ambient}$, which means $du < 0$, whereas *North-South migration* has $u^{ancestor} < u^{ambient}$, 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 expands compared to the ambient UV-R level, and thus the imbalance between actual skin pigmentation and its evolutionary optimal level, the risk of vitamin D deficiency rises above its evolutionarily 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 evolutionarily 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 $\left(u^{ancestor} - u^{ambient}\right)$, a difference which we label DIFFUV below, and life expectancy.[13] DIFFUV > 0 (i.e, $u^{ancestor} > u^{ambient}$) 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.

[13] We document below that this link is not sensitive to the inclusion of observed mortality risks from skin cancer.

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. Specifically, we are interested in constructing a variable which provides us with a measure of a population's susceptibility to vitamin D deficiency given local climatic conditions. Consequently, we need first and foremost a measure of local exposure to UV-R. Fortunately, 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 mathematical 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.[14] 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. Or, 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.[15] That is, in our analysis below we employ an average of the 1990 and 2000 observation. Figure 1 provides a map depicting the global distribution of UV-R intensity.

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. This evolutionary compromise results in paler skins, to permit vitamin D synthesis, 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

[14] *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. *Optical thickness* is a measure of the fraction of UV radiation, which 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). *Attenuation* depends on different cloud properties in complicated and partly unknown ways.

[15] 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 2 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.

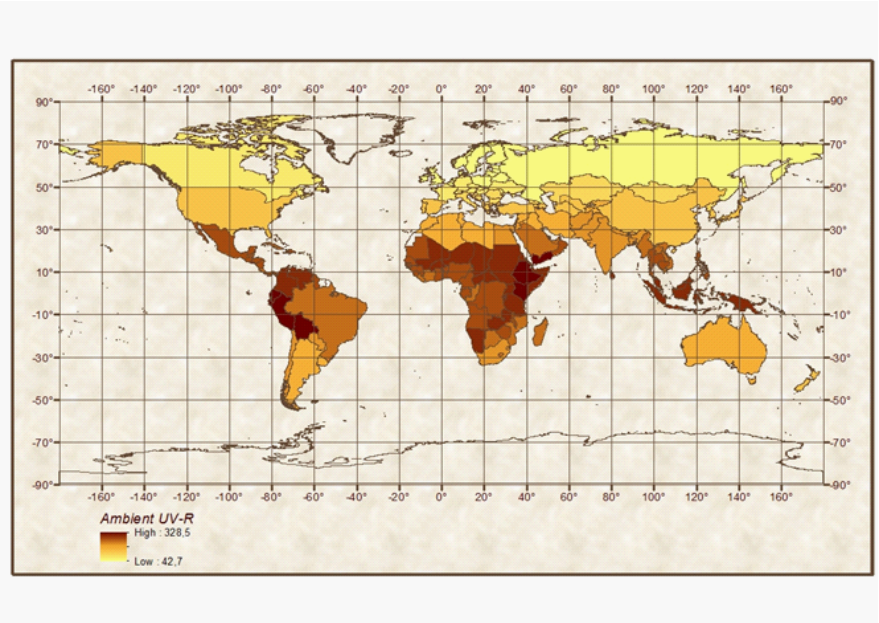

Figure 1: Global distribution of the UV-R variable. Notes: See Andersen et al. (2016) for details on construction of the UV-R index.

migration-induced perturbation.

Consequently, our approach entails ancestry adjusting 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 information on the proportion of long-term residents' ancestors of country i that is estimated to have lived in each source country in 1500. Row 43, for example, is Denmark. The row vector $\mathbf{m}_{43}$ 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

$$\text{DIFFUV}_i \equiv \mathbf{m}_i\mathbf{UV} - UV_i, \tag{1}$$

where $\mathbf{UV}$ is the N by 1 vector of UV-R levels and UV_i is UV-R in country i. DIFFUV_i then

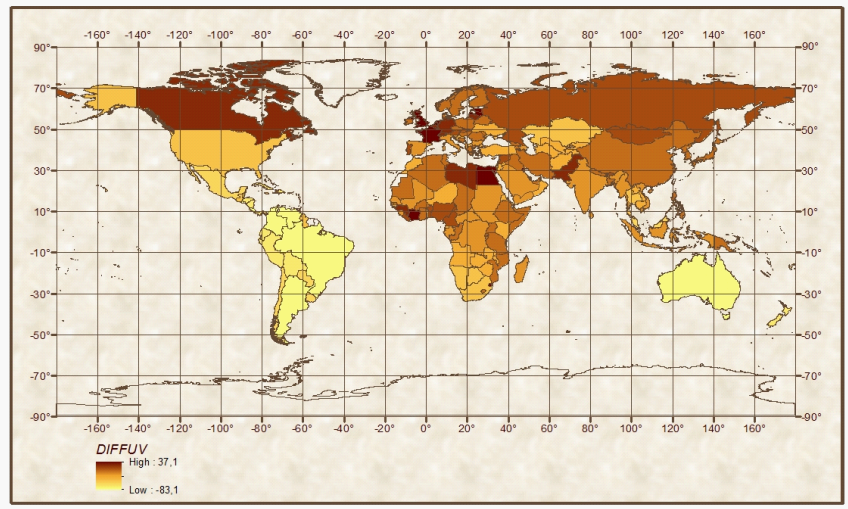

Figure 2: Global map of DIFFUV.

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 $\text{DIFFUV}_i > 0$ has a population that (on average) has a skin type more suited for living in places with stronger UV-R (as measured by $\mathbf{m}_i\mathbf{UV}$) than in the country *i* where they actually live (as measured by UV_i). That is, $\text{DIFFUV}_i > 0$ 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 $\text{DIFFUV} < 0$. 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:

$$H_i = \beta_0 + \beta_1 \text{DIFFUV}_i + \mathbf{Z}_i\gamma + \varepsilon_i, \tag{2}$$

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); $\mathbf{Z}_i$ is a vector of auxiliary controls. Our primary health indicator in the analysis below is life expectancy at birth. However, we also examine the link between DIFFUV and adult mortality, mortality for children under the age of five, and for infant mortality. In all settings the parameter of interest in equation (2) is β_1.

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, aside from a full set of (i.e., eight) regional fixed effects; identification of an influence from DIFFUV is thus sought in the *within regional variation* in health outcomes.[16] 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.[17] In the present case, however, there is an additional reason why it seems worth controlling for mean income: As economic development likely reduces the exposure time to UV-R, due to 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 2004; Ram 2006) and human capital in the form of schooling (cf., Grossman 2003).[18] Naturally, all these variables are quite likely endogenous vis-à-vis life expectancy, which we try to address by lagging

[16]The 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.

[17]See 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.

[18]We show in Section 3.3.2 that inclusion of these proximate determinants of health does not introduce a "bad controls" problem.

all variables by roughly a decade; life expectancy is measured in 2010, whereas our controls are measured in 2000 at the latest.[19] 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 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 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 enters separately in the model. As one moves from left to right we progressively add controls; column 6 include mean 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 R-squared between 0.8 and 0.9 depending on the exact specification. As can be seen, our

[19] Specifically, income inequality is measured as an average from 1960-1996; the remaining variables are measured in 2000.

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

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

TABLE 2

In Table 2 we examine a further set of contemporary determinants of health differences. 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.[20] 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).[21]

[20] These controls are generally standard in the literature on the "correlates of health"; see, e.g., Or (2000), Moore et al. (2003), and Marmot (2006). See, however, Besley and Kudamatsu (2006) on a possible democracy link.

[21] Naturally, the model already controls for per capita income. However, the latter is very likely measured with error for which reason UV-R could still emerge significant in the model.

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 lowers (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 (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.[22] 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 Cervelatti 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.

The specification in column 13 of Table 3 allows us to address any worries about the so-called "bad controls" problem (see Angrist and Pischke 2009). All the controls in the column were fixed prior to the time that DIFFUV was determined. Consequently, if we estimate the specification in column 13 on the same sample as that of column 12 of the table, we get that the

[22] This can be put on a rigorous footing. Specifically, to gauge how concerned we should be about omitted variables, we invoke the insights of Altonji et al. (2005) that selection on observables can be used to assess the likely importance of bias arising from unobservable factors. Concretely, we calculate the bias-corrected estimate derived by Oster (2016), which generalizes the work of Altonji et al. (2005). The minimal model is a regression of life expectancy on only DIFFUV, whereas the maximal model is column 12 of Table 3 (both models are estimated using the sample in column 12). We find that the bias-corrected estimate is -0.190, which is very close to the estimates found in Table 3. Consequently, our analysis appears to be robust to omitted variables.

difference in DIFFUV point estimates is statistically insignificant (not reported). The same goes for comparisons of the baseline specification (i.e., column 12 of Table 1) and column 13 of Table 13 (not reported). Consequently, a bad controls problem does not seem to plague our analysis.

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 influence the human immune system (Pearce and Cheetham 2010).

Turning to the results, Table 4 explores the link between DIFFUV and adult mortality, child (< 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 effect; 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.[23] 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 Sample splits

The fundamental hypothesis under scrutiny is that historical population movements have produced (in some regions) a mismatch between evolutionary optimal skin pigmentation and actual pigmentation. By implication, we expect to see a stronger economic impact from DIFFUV in regions that

[23] 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.

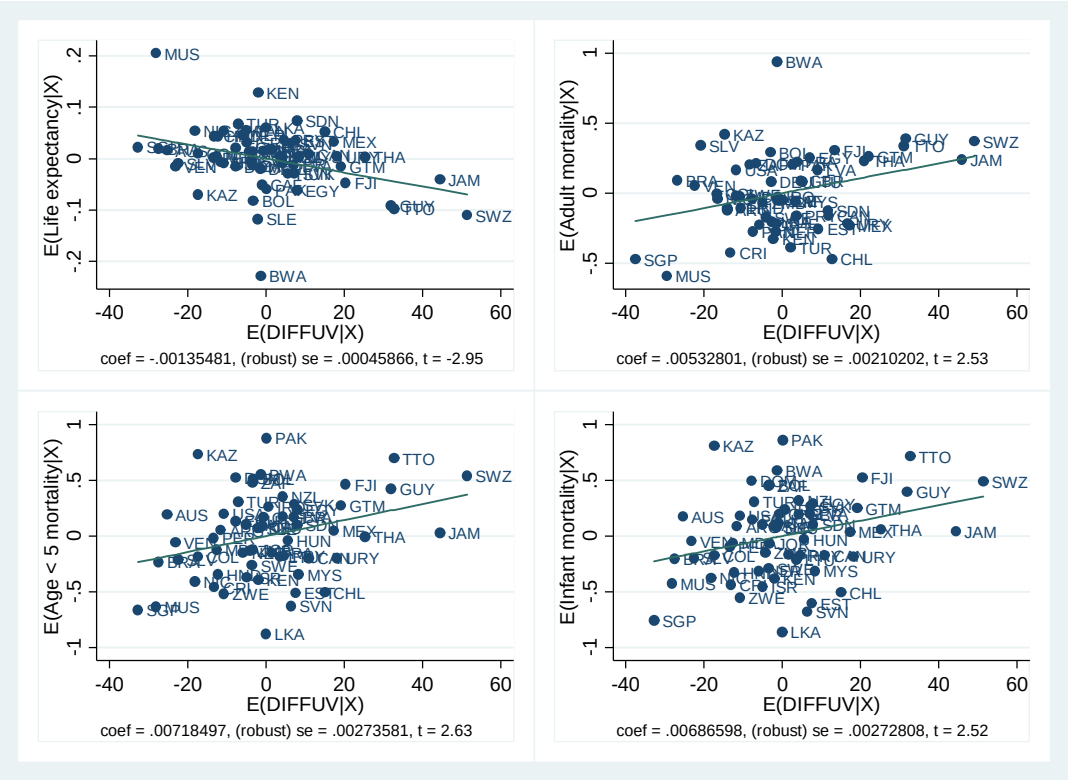

Figure 3: Partial correlation between DIFFUV and mortality. Notes: The partial correlation is conditional on the full set of baseline controls (mean income, inequality, human capital and regional fixed effects) and for a sample of countries where the percentage of native inhabitants (as of 1500 C.E.) is below 95.

have experienced more immigration. If this is not the case, our main findings likely reflect some omitted factor not accounted for in the checks conducted above. Table 5 reports the results from this consistency check.

TABLE 5

The first three columns focus on countries with limited immigration; starting from left to right, it concerns countries where *at least* 95%, 75%, and 50%, respectively, of the population can be considered native. The last three columns focuses on countries where *at most* 95%, 75%, and 50% of the population is native.

The results are quite striking. Regardless of which mortality measure we employ, a clear pattern emerges: when the immigration share increases, the impact from DIFFUV monotonically rises. In contrast, in places with insignificant immigration, we see no influence from DIFFUV. These results show quite clearly that the identification obtained above is through the reshuffling of world population after 1500 C.E.

By way of concluding this section, we provide some visual illustrations of the link between DIFFUV and the individual mortality measures. As is clear upon eyeballing Figure 3, when we focus on countries that have received a sufficiently large amount of immigration (5% or more), our risk-of-vitamin-D-deficiency measure is a clear predictor of longevity and mortality, regardless of which age group one chooses to focus on.

4 Cross-State Analysis

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

The US based 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 inevitably 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 to 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.[24] The details are laid out in Section B of the Appendix. A visual representation of DIFFUV for the US states is provided in Figure 4.

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

[24]Their Main Appendix 1.1 describes how they calculate ancestry shares.

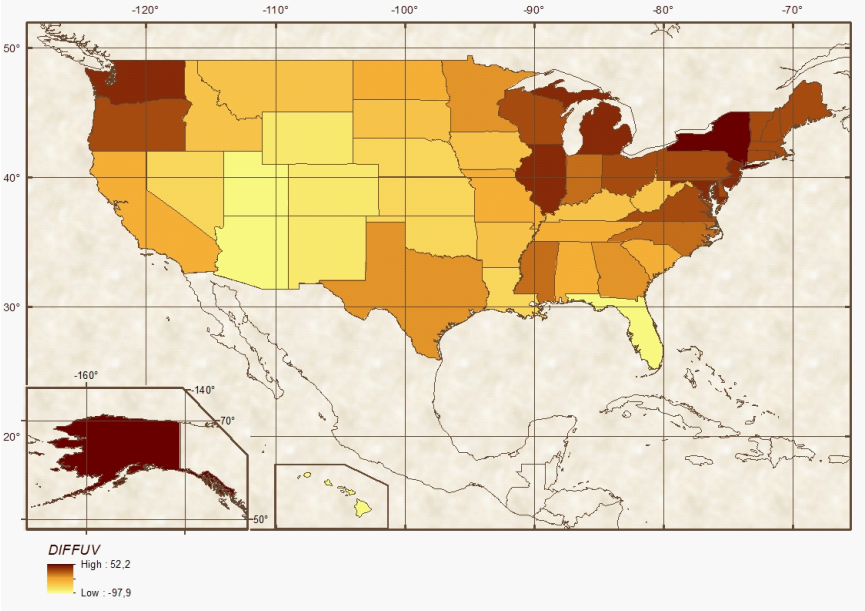

Figure 4: US map of DIFFUV.

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 4, 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 5 in Section C of the 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 Appendix.

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. However, 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.[25] Second, as our measure of human capital we employ the fraction of the population with at least a bachelor degree, rather than average years of schooling. Once again, all estimates are standardized and robust p-values are reported in parenthesis.

TABLE 6

[25]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.

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. 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 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 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) 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.[26]

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

[26]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.

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.[27]

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.

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 Noner 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

[27] 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 "staus channel" should give rise to a *monotonic* link between DIFFUV and health outcomes.

once we omit DIFFUV? Naturally, if both answers are in the affirmative then this provides further support of 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 when we add DIFFUV, EAPR turns insignificant. Naturally, if we had actual (reliable) 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 very likely is vitamin D deficiency. We thus interpret these result as a fairly strong indication 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 the answer to be in the affirmative. 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 the said 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 suggest 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 to cigaret smoking, it is worth noting that migrants also experienced marked 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.

References

[1] Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Roman Wacziarg. 2003. Fractionalization. *Journal of Economic Growth*, 8, 155-194.

[2] Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou. 2015. Ethnic Inequality. *Journal of Political Economy*, forthcoming.

[3] Altonji, Joseph G., Todd E. Edler, and Christopher R. Taber. 2005. Selection on Observed and UnobservedVariables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113, 151–184.

[4] Andersen, Thomas Barnebeck, Carl-Johan Dalgaard, and Pablo Selaya. 2016. Climate and the Emergence of Global Income Differences. *Review of Economic Studies*, 83, 1334-63.

[5] Angrist, Joshua, and Jörn-Steffen Pischke, 2009. *Mostly Harmless Econometrics*. Princeton, Princeton University Press.

[6] Baird, Sarah, Jed Friedman and Norbert Schady. 2011. Aggregate income shocks and infant mortality in the developing world. *Review of Economics and Statistics*, 93, 847-856.

[7] Barnes, Ian, et al. 2011. Ancient urbanization predicts genetic resistance to tuberculosis. *Evolution*, 65, 842-848.

[8] Besley, Timothy, and Masayuki Kudamatsu. 2006. Health and Democracy. *American Economic Review*, 92, 313-18.

[9] Bhopal, Raj S. 2014. *Migration, Ethnicity, Race, and Health in Multicultural Societies*, Second Edition. Oxford, Oxford University Press.

[10] Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2015. The impact of the Great Migration on mortality of African Americans: Evidence from the Deep South. *American Economic Review*, 105, 477-503.

[11] Calbó, Josep, David Pages, and Josep-Abel Gonzáles. 2005. Empirical studies of cloud effects on UV radiation: A review. *Reviews of Geophysics*, 43, 1-28.

[12] Camargo, Carlos A., Sunday Clark, Michael S. Kaplan, Philip Lieberman, and Robert A. Wood. 2007. Regional differences in EpiPen prescriptions in the United States: The potential role of vitamin D. *Journal of Allergy and Clinical Immunology*, 120, 131-136.

[13] Cervellati, Matteo, and Uwe Sunde. 2005. Human capital formation, life expectancy, and the process of development. *American Economic Review*, 95, 1653-1672.

[14] Clark, Damon, and Heather Royer. 2013. The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review*, 103, 2087-2120.

[15] Cockell, Charles S., and Gerda Horneck. 2001. The History of the UV Radiation Climate of the Earth - Theoretical and Space-based Observations, *Photochemistry and Photobiology*, 73, 447-51.

[16] Dalgaard, Carl-Johan, and Holger Strulik, 2014. Optimal Aging and Death: Understanding the Preston Curve. *Journal of the European Economic Association*, 12, 672-701.

[17] Deschênes, Olivier, and Enrico Moretti, 2009. Extreme weather events, mortality, and migration. *Review of Economics and Statistics*, 2009, 91, 659-681.

[18] Deaton, Angus. 2003. Health, Inequality, and Economic Development. *Journal of Economic Literature*, 41 113-58.

[19] Diamond, Jared. 2005. Geography and skin colour. *Nature*, 435, 283-284.

[20] Endocrine Society. 2011. Evaluation, Treatment, and Prevention of Vitamin D Deficiency: An Endocrine Society Clinical Practice Guideline. www.endocrine.org

[21] Fogelman, Y., Y. Rakover, and R. Luboshitsky. 1995. High prevalence of vitamin D deficiency among Ethiopian women immigrants to Israel: excerbation during pregnancy and lactation. *Israeli Journal of Medical Science*, 31, 221-224.

[22] Galor, Oded, and Omer Moav. The Neolithic Revolution and Contemporary Variations in Life Expectancy. Working Paper 2007-14, Brown University, Department of Economics, 2007.

[23] Garland, Cedric F., Frank C. Garland, Edward D. Gorham, Martin Lipkin, Harold Newmark, Sharif B. Mohr, and Michael F. Holick. 2006. The Role of Vitamin D in Cancer Prevention. *American Journal of Public Health*, 96, 252-261.

[24] Ginde, Adit A., Mark C. Liu, and Carlos A Camargo. 2009. Demographic Differences and Trends of Vitamin D Insufficiency in the US Population, 1988-2008. *Archives of Internal Medicine*, 169, 626-632.

[25] Grossman, Michael. 2003. Education and nonmarket outcomes. In: Hanushek, E., Welch, F. (Eds.), *Handbook of the Economics of Education*. North-Holland, Elsevier Science, Amsterdam.

[26] Hansen, Casper Worm. 2013. The Diffusion of Health Technologies: Cultural and Biological Divergence, *European Economic Review*, 64, 21-34.

[27] Henderson, J., M. Dunnigan, W. McIntosh, A. Abdul-Motaal, G. Gettinby, and B. Glekin. 1987. The importance of limited exposure to ultraviolet radiation and dietary factors in the aetiology of Asian rickets: a risk factor model. *Quarterly Journal of Medicine*, 63, 413-425.

[28] Holick, Michael F. 2007. Vitamin D deficiency. *New England Journal of Medicine*, 357: 266-81.

[29] Jablonski, Nina G., and George Chaplin. 2000. The evolution of human skin coloration, *Journal of Human Evolution*, 39, 57-106.

[30] Kulie, Teresa, Amy Groff, Jackie Redmer, Jennie Hounshell, and Sarina Schrager. 2009. Vitamin D: An Evidence-Based Review. *Journal of the American Board of Family Medicine*, 22, 698-706.

[31] Marmot, Michael. 2006. Health in an unequal world. *The Lancet*, 368, 2081-2094.

[32] Melamed, M., E. Michos, W. Post., B. Astor. 2008. 25-hydrxyvitamin D levels and the risk of mortality in the general population, *Archives of Internal Medicine*, 168, 1340-1349.

[33] Moore, Melinda, Philip Gould, and Barbara S. Keary. 2003. Global urbanization and impact on health, *International Journal of Hygiene and Environmental Health*, 2006, 269-278

[34] Mullins, R., and C. Camargo. 2012. Latitude, sunlight, vitamin D, and childhood food allergy/anaphylaxis. *Current Allergy and Asthma Reports*, 12, 64-71.

[35] Nwaru, B., S. Ahonen, M. Kaila, M. Erkklola, A. Haapala, C. Kronberg-Kippilä, R. Veijola, J. Ilonen, O. Simell, M. Knip, and S. Virtanen. 2010. Maternal diet during pregnancy and allergic sensitization in the offspring by 5 yrs of age: a prospective cohort study. *Pediatric Allergy and Immunology*, 21, 29-37.

[36] OECD 2015. International Migration Outlook 2015. OECD Publishing, Paris.

[37] Oster, Emily. 2016. Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business Economics and Statistics*, forthcoming.

[38] Or, Zeynep. 2000. Determinants of Health Outcomes in Industrialised Countries: A Pooled, Cross Country, Time-Series Analysis. *OECD Economic Studies*, 30, 53-77.

[39] Pearce, Simon H. S., and Tim D. Cheetham. 2010. Diagnosis and management of vitamin D deficiency, *British Medical Journal*, 340, 142-147

[40] Peroni, Diego G., and Attilio L. Boner. 2013. Food Allergy: The Perspectives of Prevention Using Vitamin D. *Current Opinion in Allergy and Clinical Immunology*, 13, 287-292.

[41] Preston, Samuel H. The changing relation between mortality and level of economic development. *Population Studies*, 29, 231-248.

[42] Putterman, Louis, and David N. Weil. 2010. Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality. *Quarterly Journal of Economics*, 125, 1627-1682.

[43] Ram, Rati. 2006. Further examination of the cross-country association between income inequality and population health. *Social Science & Medicine*, 62, 779-791.

[44] Sachs, Jeffrey D. 2003. Institutions Don't Rule: Direct Effects of Geography on Per Capita Income. NBER Working Paper, 9490, National Bureau of Economic Research, Cambridge, Ma.

[45] Soares, Rodrigo R, 2007. On the determinants of mortality reductions in the developing world. *Population and Development Review*, 33, 247-287.

[46] Sharief S., S. Jariwala, J. Kumar, P. Muntner, and M. Melamed. 2011. Vitamin D levels and food and environmental allergies in the United States: results from the National Health and Nutrition Examination Survey. *Journal of Allergy and Clinical Immunology*, 127, 1195-1202.

[47] United Nations, 2014. Population Facts. 2013/3 Rev.1, April.

[48] United Nations High Commissioner for Refugees, 2016. Global Trends–Forced Displacement in 2015. Geneva.

[49] Williams, D. R., 1999. Race, socioeconomic status, and health: the added effects of racism and discrimination. *Annals of the New York Academy of Sciences*, 896, 173-188.

A Summary Statistics and Data Sources

A.1 Cross-country

TABLE B1

A.2 US states

TABLE B2

B Constructing DIFFUV across US states

In order to generate DIFFUV across US states we need to construct a US state level migration matrix as well as to calculate average UV-R levels for each state. To construct the migration matrix we use the same data and methodology as Putterman and Weil (2010) (henceforth PW). Concretely, we use the ancestor variables from the IPUMS 2000 and apply the appropriate weights to obtain state level representativeness. Next, for each state we sum the (weighted) number of persons reporting each ancestry.[28] We then group people that report similar ancestry like 'Spanish' and 'Spanish American' and make adjustments to match ancestries to present day country borders. For example, people reporting 'Yugoslavian' ancestry are allocated to the five countries making up the former Yugoslavia based on their relative population shares in 2000.[29]

To trace the roots all the way back to 1500 the majority that report an American country as ancestry need to be traced further back to their 'Old World' roots. In order to do so we use the world migration matrix and apply the ancestries of those respective countries, and reallocate relative to

[28] People are allowed to report up to two ancestries. We follow PW and put equal weight on both ancestries when reported. Approximately 80% report one ancestry while 22 % report two ancestries. This means that in some states the response rate will be above 100%. To compensate for this we normalize each state level ancestry share by dividing by the state's response rate.

[29] People reporting either 'Arab' or 'Arabic' or 'Other Arab' are allocated in the same relative proportions to those declaring Egypt, Syria and Lebanon as countries of ancestry. People reporting 'Czechoslovakian' ancestry are allocated in same relative proportions to those declaring Czech and Slovak ancestry. People reporting 'Scandinavian' ancestry are allocated in same relative proportions to those declaring Sweden, Denmark, Finland and Norway as countries of ancestry. Those reporting 'Hispanic' ancestry are allocated in same relative proportions to those declaring Mexico, Puerto Rico, Guatemala, El Salvador, The Dominican Republic, Cuba, Columbia, Ecuador, Peru, Honduras and Nicaragua as countries of ancestry. People reporting 'Latin American' ancestry are allocated similarly except that the Spanish-speaking Caribbean countries (Puerto Rico, Dominican Republic and Cuba) and Guatemala are left out. People reporting 'Asian' ancestry are allocated in same relative proportions to those declaring China, Japan, S. Korea, India and Pakistan as countries of ancestry. People reporting 'Irish' ancestry are allocated in relative proportions to the population sizes (as of 2000) of the two parts making up the island (U.K. and Ireland). People reporting 'West Indian' ancestry are allocated in same relative proportions to those declaring Jamaica and Trinidad and Tobago as countries of ancestry.

the ancestry proportions of the present-day populations.[30] People with 'African' ancestry are allocated to countries based on US port of disembarkation of slaves.[31] These data come from "The Trans-Atlantic Slave Trade" database, which identifies ports of embarkation and disembarkation. PW use these data and present-day geographical information of the embarkation regions to estimate the slaves' countries of origin.[32]

We finally allocate people reporting European or North American ancestries. First people with 'White' or 'European' ancestry are allocated to European countries in the same proportions as those reporting specific European countries of ancestry, while people specifying a certain region (e.g. 'Western Europe') are allocated similarly between countries in that specific region. People reporting 'American Indian' or 'Hawaiian' ancestry are assumed to be Native Americans[33] while those reporting 'US', 'North American', 'American', or 'Other ancestry' are assumed not to be Native Americans and are allocated proportionally to the rest of the world.

After having allocated all people to specific countries, we follow PW and ignore ancestries with only few people. While they ignore ancestries representing less than 0.1% of the US population, we work with a more disaggregated level and hence choose only to ignore ancestries representing less than 0.01% of a state population.[34]

To calculate state level UV-R we follow Andersen et al. (2016) and calculate the yearly average UV-R for each grid cell based on NASA's daily local noon observations from 1990 and 2000. We weight these by population using data from NASA's Socioeconomic Data and Applications Center (SEDAC) Gridded Population of the World (GPW, v3) dataset aggregated to the same resolution as the UV-R data. In total we use more than 600,000 observations to calculate state level averages.

With these data in hand we can calculate DIFFUV for each state in accordance with equation (1).

C Map of Ancestor UV-R across US States

[30] Doing so, we assume—like PW—that people have come to the US in same relative proportions to the ancestries of each country.

[31] People reporting African-American ancestry, 80% are assumed to be African, 19% are assumed to be European, and 1% are assumed to be native Americans

[32] See Table 3 of their Main Appendix 1.1, Section II-3.

[33] As a rough approximation we assume that Native Americans live in the same state as their ancestors.

[34] We could have followed PW more closely and deleted all ancestries (in all states) that on the national level represented less than 0.1% of the entire US population, but given the geographical clustering of some ancestries that would have removed (on the state level) significant shares of the population.

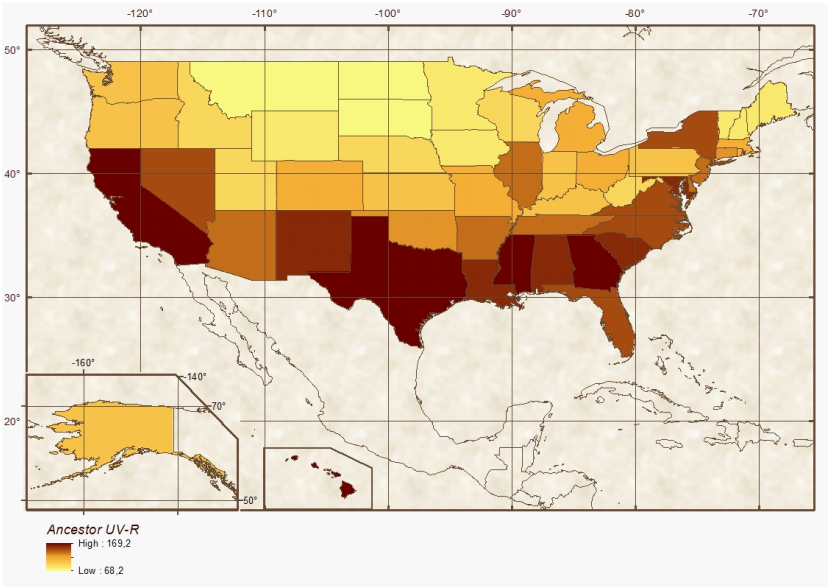

Figure 5: US map of ancestor UV-R.

Table 1: The Augmented Preston Curve

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Life expectancy (log)											
DIFFUV							-0.242	-0.297	-0.215	-0.166	-0.160	-0.180
							(0.000)	(0.000)	(0.002)	(0.017)	(0.027)	(0.008)
UV-R (ancestor)	-1.510	-0.979	-0.697	-0.485	-0.475	-0.553						
	(0.000)	(0.000)	(0.002)	(0.024)	(0.026)	(0.007)						
UV-R (ambient)	0.875	0.890	0.704	0.572	0.494	0.553						
	(0.000)	(0.000)	(0.003)	(0.018)	(0.035)	(0.011)						
GDP per capita (log)			0.282	0.292	0.321	0.217			0.290	0.275	0.323	0.227
			(0.000)	(0.009)	(0.007)	(0.183)			(0.000)	(0.008)	(0.006)	(0.156)
Inequality (Gini)				-0.134		-0.287				-0.130		-0.285
				(0.031)		(0.000)				(0.034)		(0.000)
Human capital (years of schooling)					0.041	-0.030					0.046	-0.013
					(0.583)	(0.767)					(0.501)	(0.890)
Regional fixed effects	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Observations	157	157	157	120	119	98	157	157	157	120	119	98
R-squared	0.52	0.81	0.83	0.83	0.84	0.86	0.06	0.80	0.83	0.83	0.84	0.86
H0: UV (anc) + UV (amb) = 0 (p-value)	0.000	0.066	0.625	0.592	0.862	0.633						

Robust p-values in parentheses.

Table 2. The Augmented Preston Curve: Contemporary Structural Characteristics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Life expectancy (log)								
DIFFUV	-0.165	-0.181	-0.169	-0.192	-0.194	-0.182	-0.183	-0.192	-0.199
	(0.040)	(0.009)	(0.015)	(0.003)	(0.010)	(0.008)	(0.007)	(0.036)	(0.017)
Skin cancer (death rate)	0.043							0.028	0.114
	(0.540)							(0.689)	(0.064)
Health expenditure (% of GDP)		0.013						0.017	-0.018
		(0.806)						(0.775)	(0.724)
Alcohol consumption			-0.096					-0.137	-0.004
			(0.251)					(0.141)	(0.942)
Air pollution				0.113				0.127	0.104
				(0.156)				(0.104)	(0.045)
Urbanization					-0.056			-0.083	0.084
					(0.482)			(0.413)	(0.206)
Female participation rate						0.027		0.034	0.057
						(0.693)		(0.646)	(0.446)
Democracy (Polity IV)							0.025	0.017	0.081
							(0.714)	(0.810)	(0.186)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	98	96	98	98	98	98	98	96	139
R-squared	0.86	0.85	0.86	0.86	0.86	0.86	0.86	0.87	0.83

Robust p-values in parentheses. Baseline controls are GDP per capita, income inequality, and schooling.

Table 3. The Augmented Preston Curve: Geographical and Historical Structural Characteristics

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Life expectancy (log)												
DIFFUV	-0.206	-0.284	-0.140	-0.115	-0.178	-0.190	-0.182	-0.207	-0.190	-0.183	-0.144	-0.338	-0.379
	(0.003)	(0.042)	(0.039)	(0.097)	(0.011)	(0.012)	(0.006)	(0.006)	(0.010)	(0.008)	(0.057)	(0.068)	(0.006)
UV-R (ambient)	-0.108											-0.640	-0.610
	(0.320)											(0.167)	(0.003)
DIFFABSLAT		-0.101										-0.150	-0.174
		(0.289)										(0.393)	(0.139)
Years since Neolitic (ancestor adj.)			0.120									0.035	0.103
			(0.093)									(0.637)	(0.058)
Genetic distance (USA)				-0.170								-0.035	0.006
				(0.031)								(0.694)	(0.945)
Malaria ecology					0.015							0.078	0.072
					(0.818)							(0.361)	(0.337)
HLA heterozygosity (log)						0.005						-0.051	-0.049
						(0.946)						(0.680)	(0.538)
Tropical							-0.006					-0.091	-0.142
							(0.929)					(0.320)	(0.105)
Year of fertility decline								-0.065				0.015	-0.122
								(0.346)				(0.866)	(0.076)
Absolute latitude									0.039			-0.546	-0.742
									(0.691)			(0.144)	(0.000)
Ethnic fractionalization										-0.012		-0.016	-0.030
										(0.810)		(0.816)	(0.560)
Ethnic inequality											0.002	-0.025	-0.055
											(0.965)	(0.713)	(0.308)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	98	98	97	96	98	92	98	79	98	98	94	68	104
R-squared	0.86	0.86	0.86	0.87	0.86	0.86	0.86	0.87	0.86	0.86	0.87	0.90	0.88

Robust p-values in parentheses. Baseline controls are GDP per capita, income inequality, and schooling.

Table 4. The Augmented Preston Curve: Alternative Mortality Measures

VARIABLES	(1) Adult mortality (log)	(2) Age < 5 mortality (log)	(3) Infant mortality (log)	(4) Adult mortality (log)	(5) Age < 5 mortality (log)	(6) Infant mortality (log)	(7) Adult mortality (log)	(8) Age < 5 mortality (log)	(9) Infant mortality (log)
DIFFUV	0.267	0.334	0.313	0.138	0.155	0.152			
	(0.003)	(0.000)	(0.000)	(0.122)	(0.008)	(0.022)			
UV-R (ancestor)							0.581	0.557	0.553
							(0.018)	(0.002)	(0.006)
UV-R (ambient)							-0.348	-0.423	-0.410
							(0.173)	(0.012)	(0.035)
GDP per capita (log)				-0.319	-0.530	-0.539	-0.254	-0.488	-0.495
				(0.093)	(0.000)	(0.000)	(0.176)	(0.000)	(0.000)
Inequality (Gini)				0.311	0.111	0.120	0.317	0.115	0.125
				(0.000)	(0.016)	(0.015)	(0.000)	(0.015)	(0.013)
Human capital (years of schooling)				0.084	-0.264	-0.269	0.174	-0.206	-0.207
				(0.584)	(0.002)	(0.004)	(0.288)	(0.064)	(0.078)
Regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	143	143	143	86	86	86	86	86	86
R-squared	0.72	0.72	0.69	0.78	0.90	0.88	0.79	0.90	0.88
H0: UV (anc) + UV (amb) = 0 (p-value)							0.031	0.164	0.167

Robust p-value in parentheses.

Table 5: Augmented Preston Curve: Sample Splits

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Life expectancy (log)					
DIFFUV	-0.080	-0.020	-0.024	-0.270	-0.311	-0.387
	(0.277)	(0.772)	(0.775)	(0.005)	(0.006)	(0.032)
Observations	38	57	72	60	41	25
R-squared	0.90	0.90	0.89	0.85	0.82	0.86
	Adult mortality (log)					
DIFFUV	0.133	-0.102	-0.071	0.271	0.315	0.357
	(0.166)	(0.242)	(0.620)	(0.015)	(0.014)	(0.053)
Observations	32	48	63	54	38	22
R-squared	0.88	0.87	0.84	0.79	0.71	0.84
	Age < 5 mortality (log)					
DIFFUV	0.028	-0.010	-0.002	0.205	0.242	0.281
	(0.428)	(0.872)	(0.979)	(0.012)	(0.010)	(0.018)
Observations	38	57	72	60	41	25
R-squared	0.96	0.93	0.92	0.88	0.87	0.89
	Infant mortality (log)					
DIFFUV	0.024	-0.018	-0.011	0.208	0.240	0.255
	(0.529)	(0.794)	(0.915)	(0.015)	(0.012)	(0.029)
Observations	38	57	72	60	41	25
R-squared	0.95	0.92	0.91	0.86	0.84	0.88
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
% of population native	> 95%	> 75%	> 50%	< 95%	< 75%	< 50%

Robust p-values in parentheses. Baseline controls are GDP per capita, income inequality, and schooling.

Table 6. The Augmented Preston Curve across US States

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Life expectancy (log)											
UV-R (ambient)	0.045	0.254	0.478	0.237	0.506	0.470						
	(0.818)	(0.027)	(0.001)	(0.106)	(0.000)	(0.004)						
UV-R (ancestor)	-0.361	-0.096	-0.376	-0.041	-0.492	-0.343						
	(0.015)	(0.593)	(0.045)	(0.881)	(0.006)	(0.078)						
DIFFUV							-0.055	-0.201	-0.367	-0.206	-0.368	-0.363
							(0.679)	(0.025)	(0.001)	(0.028)	(0.001)	(0.003)
Income per capita (log)			0.410			0.022			0.389			0.020
			(0.001)			(0.869)			(0.004)			(0.881)
Income inequality (Gini)				-0.057		-0.187				0.021		-0.205
				(0.710)		(0.173)				(0.829)		(0.066)
Human capital					0.496	0.514					0.422	0.511
					(0.000)	(0.002)					(0.007)	(0.002)
Regional fixed effects	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Observations	51	51	51	51	51	51	51	51	51	51	51	51
R-squared	0.11	0.71	0.78	0.71	0.81	0.82	0.00	0.70	0.77	0.70	0.80	0.82
H0: UV (anc) + UV (amb) = 0 (p-value)	0.026	0.584	0.647	0.561	0.164	0.816						

Robust p-values in parentheses.

Table 7. Vitamin D Deficiency and DIFFUV: The Mechanism

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Epinephrine autoinjector prescription rate (EAPR)					Life expectancy (log)	
DIFFUV	0.408	0.341	0.365	0.422	0.351		-0.342
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		(0.006)
EAPR						-0.335	-0.060
						(0.080)	(0.694)
Income per capita (log)		0.155			0.140	-0.036	0.028
		(0.075)			(0.294)	(0.817)	(0.835)
Human Capital			0.108		0.064	0.526	0.514
			(0.192)		(0.643)	(0.005)	(0.002)
Income inequality (Gini)				-0.058	-0.120	-0.265	-0.212
				(0.570)	(0.279)	(0.047)	(0.059)
Regional fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51	51	51	51	51	51	51
R-squared	0.82	0.83	0.83	0.82	0.84	0.78	0.82

Robust p-values in parentheses.

Table B1.

	# Obs	Mean	Std. dev.	Min	Max	Description	Source
Life expectancy (log)	157	4.22	0.15	3.80	4.42	Life expectancy at birth, total years.	World Development Indicators.
Adult mortality (log)	143	5.13	0.59	4.04	6.58	Adult mortality rate is the probability (per 1000) of dying between the ages of 15 and 60.	World Development Indicators.
Age < 5 mortality (log)	157	3.12	1.19	0.88	5.34	Age < 5 mortality rate is the probability (per 1000) that a newborn baby will die before reaching age five.	World Development Indicators.
Infant mortality (log)	157	2.87	1.10	0.64	4.70	Infant mortality rate is the number of infants (per 1000) dying before reaching one year of age.	World Development Indicators.
DIFFUV	157	-11.12	22.70	-89.31	37.14	See equation (1) in main text.	See main text.
UV-R (ancestor)	157	183.20	72.10	42.66	305.14	Ultraviolet radiation, ancestry adjusted.	See main text.
UV-R (ambient)	157	194.40	76.18	42.66	328.53	Ultraviolet radiation, ambient.	Andersen et al. (2015).
GDP per capita (log)	157	8.40	1.38	4.76	11.21	Real GDP per capita (chain index).	Penn World Tables 7.0.
Inequality (Gini)	120	42.56	9.16	23.97	67.46	Gini coefficint, average 1960-1996.	Easterly (2008).
Human capital	119	4.79	2.89	0.41	10.86	Years of schooling.	World Development Indicators.
Skin cancer	157	0.19	0.23	0.00	1.55	Death due to skin cancer (% of total deaths).	World Health Organization's Global Health Observatory Data Repository
Health expenditure (% of GDP)	145	5.73	2.02	1.40	13.40	Health expenditure (% of GDP).	World Development Indicators.
Alcohol consumption	157	6.40	4.22	0.10	17.50	Total per capita (+15) alcohol consumption (liters of pure alcohol)	World Development Indicators.
Air pollution	157	68.19	39.05	0.00	100.00	Percent of population exposed to ambient concentrations of PM2.5 that exceed the WHO guideline value.	World Development Indicators.
Urbanization	153	54.12	23.48	5.90	100.00	Urban population (% of total).	World Development Indicators.
Female participation rate	157	52.66	14.23	19.10	82.60	Female percentage of labor force.	World Development Indicators.
Democracy	148	5.43	3.92	0.00	10.00	DEMOC (institutionalized democracy) index from Polity IV.	Polity IV Project.
DIFFABSLAT	164	-2.23	5.52	-29.35	13.10	See equation (1) in main text and substitute absolute latitude for UV-R.	See main text.
Years since Neolitic (ancestry adj.)	153	5407.00	2112.38	1356.99	10400.00	Time elapsed since Neolithic revolution (ancestry adjusted).	Ashraf and Galor (2013).
Genetic distance (USA)	145	0.09	0.08	0.00	0.23	Genetic distance to USA.	Spolaore and Wacziarg (2009).
Malaria ecology	155	3.78	6.69	0.00	31.55	Malaria ecology (population weighted)	Sachs (2003).
HLA heterozygosity (log)	146	-1.15	0.07	-1.45	-1.04	The expected heterozygosity (genetic diversity) of a country's contemporary population.	Alesina et al. (2015).
Tropical (%)	157	35.04	43.55	0.00	100.00	Tropical area (% of total).	Nunn and Puga (2012).
Year of fertility decline	126	1966.11	29.89	1865	2000	Year the country initiated the fertility decline.	
Absolute latitude	157	26.90	17.18	0.53	64.48	Absolute latitude.	Ashraf and Galor (2013).
Ethnic fractionalization	157	0.46	0.25	0.00	0.93	Ethnic fractionalization	Alesina et al. (2003).
Ethnic inequality	147	1.95	1.01	0.00	4.52	Ethnic-linguistic inequality	Alesina et al. (2015).

Table B2.

	# Obs	Mean	Std. dev.	Min	Max	Description	Source
Life expectancy (log)	51	4.36	0.02	4.31	4.40	Life expectancy at birth, total years.	Measure of America project of the Social Science Research Council.
EAPR	51	5.70	2.35	2.70	11.80	Number of EpiPen prescriptions per 1000 persons.	Carmargo et al. (2007).
UV-R (ambient)	51	126.49	33.89	48.33	239.45	See Appendix C.	See Appendix C.
UV-R (ancestor)	51	103.53	22.60	68.21	169.20	See Appendix C.	See Appendix C.
DIFFUV	51					See equation (1) in main text.	See Appendix C.
Income per capita (log)	51	10.15	0.16	9.86	10.63	Income per capita (log).	American Community Survey.
Income inequality	51					Gini coefficient.	US Census Bureau and the American Community Survey.
Human Capital	51	27.59	5.56	17.30	48.50	Percentage with bachelor degree or more.	US Census Bureau