**Geographical pattern of minerals and its association with health disparities in the USA**

Bingjie Qu, Shiqiang Wu, Peng Zhao, Zheng Feei Ma, Royston Goodacre, Linxi Yuan, Ying Chen

Bingjie Qu, Wisdom Lake Academy of Pharmacy, Xi’an Jiaotong-Liverpool University, China; Institute of Systems, Molecular & Integrative Biology, University of Liverpool, Liverpool, UK.

Shiqiang Wu, Department of Health and Environmental Sciences, Xi’an Jiaotong-Liverpool University, Suzhou, China.

Peng Zhao, Department of Health and Environmental Sciences, Xi’an Jiaotong-Liverpool University, Suzhou, China.

Zheng Feei Ma, Centre for Public Health and Wellbeing, School of Health and Social Wellbeing, College of Health, Science and Society, University of the West of England, Bristol, UK.

Royston Goodacre, Centre for Metabolomics Research, Department of Biochemistry and Systems Biology, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, UK.

Linxi Yuan, Department of Health and Environmental Sciences, Xi’an Jiaotong-Liverpool University, Suzhou, China.

Ying Chen, Wisdom Lake Academy of Pharmacy, Xi’an Jiaotong-Liverpool University, Suzhou, China.

Correspondence: Ying.Chen01@xjtlu.edu.cn (Ying Chen), ORCID: 0000-0002-5919-743X.

Acknowledgements:

The authors thank the other colleagues at the Wisdom Lake Academy of Pharmacy, Xi’an Jiaotong-Liverpool University for their academic and administrative support. YC would like to express his special appreciation for the arrival of his baby daughter, with whom the process of manuscript writing has become a particularly memorable and enjoyable moment.

**Abstract**

This study aimed to determine the common latent patterns of geographical distribution of health-related minerals across the USA, and to evaluate the real-world cumulative effects of these patterns on overall population health. It was an ecological study using county-level data (3080 contiguous counties) on the concentrations of 14 minerals (i.e. aluminum, arsenic, calcium, copper, iron, lead, magnesium, manganese, mercury, phosphorus, selenium, sodium, titanium, zinc) in stream sediments (or surface soils), and the measurements of overall health including life expectancy at birth, age-specific mortality risks and cause-specific (summarized by 21 mutually exclusive groups) mortality rates. Latent class analysis (LCA) was employed to identify the common clusters of life expectancy-related minerals based on their concentration characteristics. Multivariate linear regression analyses were then conducted to examine the relationship between the LCA derived clusters and the health measurements, with adjustment for potential confounding factors. Five minerals (i.e. arsenic, calcium, selenium, sodium and zinc) were associated with life expectancy, and were analyzed in LCA. Three clusters were determined across the USA, the ‘common’ (*n*=2056, 66.8%), ‘infertile’ (*n*=739, 24.0%) and ‘plentiful’ (*n*=285, 9.3%) clusters. Residents in counties with the ‘infertile’ profile was associated with the shortest life expectancy, highest mortality risks at all ages, and highest mortality rates for many reasons including the top five leading causes of death: cardiovascular diseases, neoplasms, neurological disorders, chronic respiratory conditions, and diabetes, urogenital, blood and endocrine diseases. Results remained statistically significant after confounding adjustment. Our study brings novel perspectives regarding environmental geochemistry to explain health disparities in the USA.

**Key words:** Geochemistry, life expectancy, age-specific mortality risk, cause-specific mortality rate, latent class analysis, ecological study

**Background**

At least since the famous article ‘Air, Water and Places’, written by Hippocrates in 400 BCE, human society has been aware of the important connections between local environment and health. Living organisms, including humans, depend on minerals, and the amount of minerals in plans, animals and humans are strongly influenced by geology. Exposure to minerals from the local geology can occur through a variety of pathways such as food, drinking water and air (“Medical Geology and the Soil Health-Human Health Nexus” 2019). Minerals can be beneficial or harmful to life, and their concentrations and abundances affect health in different ways. Seven ‘macro minerals’, including calcium (Ca), chlorine (Cl), magnesium (Mg), phosphorus (P), potassium (K), sodium (Na), and sulfur (S), and eight ‘trace elements’, including cobalt (Co), copper (Cu), iodine (I), iron (Fe), manganese (Mn), molybdenum (Mo), selenium (Se) and zinc (Zn), are defined as essential for electrolyte balance, structural and functional roles in human body, while for ‘heavy metals’, such as arsenic (As), cadmium (Cd) and lead (Pb), the effects are mainly toxic (Zoroddu et al. 2019) (Farag et al. 2021).

Minerals in the environment have been shown to impact the health of local populations (Dinh et al. 2018). Low concentration of Zn in soil is one of the major factors associated with Zn deficiency in local crops and human populations (Alloway 2009) (Gashu et al. 2021). A deficiency of Zn can lead to diffuse adverse effects on several organ systems, including the epidermal, gastrointestinal tract, central nervous, immune, skeletal and reproductive systems (Hambidge 2000). The concentration of Se in local soil determines to a great extent the level of Se intake of the local population (Hasan 2021), and is closely associated with the occurrences of Keshan and Kashin-Beck diseases (Zhang et al. 2019) (Zha et al. 2022). Other notable examples from previous epidemiological studies included that the exposure to elevated amount of Mn in drinking water during childhood was associated with increased risks of attention-deficit hyperactivity disorder problems (Schullehner et al. 2020) (S. M. Rahman et al. 2017), and that the environmental Cd was found to be associated with mortality from influenza and pneumonia in adult populations (Park et al. 2020).

Individual minerals, in association with particular health outcomes, have been studied in more detail in the literature. However, a more systematic understanding of spatial distribution of multiple health-related minerals across a large geographical area and its collective effects on health disparities is lacking. In the current study we first aimed to identify whether there were geographical associations between the concentrations of a wide variety of minerals in stream sediments and surface soils, and the life expectancy of local residents, using the national data of USA. Second, we aimed to determine the common latent patterns of geographical distribution of these life expectancy-associated minerals, using cluster analysis. Finally, using these derived patterns, we aimed to evaluate the real-world cumulative effects of minerals on the measurements of overall health including life expectancy at birth, age-specific mortality risks and cause-specific mortality rates.

**Materials and Methods**

*Setting and design*

This was designed as an ecological study where the sample unit was the counties (or county/state equivalents) of USA, with two main areas of county-level data: 1) mineral concentration data, and 2) statistics providing the overall health evaluation of local population (including life expectancy at birth, age-specific mortality risk and cause-specific mortality rate). In addition, characteristics of population, socioeconomics, healthcare service, and residential environment and location by county were also included for potential confounding adjustment.

*Database*

Mineral concentration data were obtained from the latest version of the National Geochemical Survey (updated in September 2008). The survey was a national-scale geochemical analysis in the USA, from existing data, reanalysis of existing samples, and new sampling, conducted by the United States Geological Survey in collaboration with other federal and state government agencies, industry and academia (Smith et al. 2013). The survey aims to produce a body of geochemical data based primarily on stream sediments, analyzed using a consistent set of methods in order to provide the possible maximum level of internal consistency (Smith et al. 2013). The goal of the National Geochemical Survey is to analyze at least one stream-sediment sample in every 289 km2 area by a single set of analytical methods across the entire nation, with other solid sample media substituted where necessary. Samples were from approximately a depth of 20 cm from the surface. For the county level concentrations, data were available in the conterminous states for the following 14 minerals (Al, As, Ca, Cu, Fe, P, Pb, Mg, Mn, Hg, Na, Se, Ti, Zn), which were considered to be the elements of importance to environment research (Smith et al. 2013). As, Hg and Se concentrations were determined by hydride-generation atomic absorption spectrometry (HG-AAS), while the others were measured using inductively coupled plasma-atomic emission spectrometry (ICP-AES) (Smith et al. 2013). Concentrations were reported as mg/kg (mg of substance per kg of solid sample).

The database contained chemical analyses for more than 70,000 samples covering the majority of counties or county equivalents (3086/3140, 98.3% of the total counties). Geochemical data were only collected in the contiguous states, so Alaska and Hawaii were not included. Data were also not available on the following 25 counties: St. Mary (in Louisiana), Nantucket (in Massachusetts), Keweenaw (in Michigan), Hudson (in New Jersey), Major and Woodward (in Oklahoma), Camp, Delta, Franklin, Gregg, Hansford, Hopkins, Kinney, Lipscomb, Loving, Maverick, Morris, Ochiltree, Rains, Smith, Titus, Upshur and Wood (in Texas), and Lexington and Manassas Park City (in Virginia).

Population health data, including life expectancy at birth and age-specific mortality risk, were obtained from the Institute for Health Metrics and Evaluation (IHME). In the research by IHME, small area estimation methods were used to construct annual life tables, and calculate age-specific mortality risks. De-identified mortality data from the National Center for Health Statistics (NCHS) and population counts from the census bureau, NCHS, and the Human Mortality Database were used for the analysis. This dataset provides estimates of life expectancy at birth and age-specific mortality risk for each county or county equivalent from 1980 to 2014 (in every 5 years except for 2014), as well as the changes in life expectancy and age-specific mortality risk during this period. Data on age-specific mortality risk was provided in the following age categories 0-5, 5-25, 25-45, 45-65, and 65-85 (Dwyer-Lindgren et al. 2017).

IHME research applied a novel methodology to death registration data from the National Vital Statistics System in order to estimate annual mortality rates for 21 mutually exclusive causes of death. This dataset provides estimates for cause-specific age-standardized mortality rates at the county level for each county or county equivalent from 1980 to 2014 (in every 5 years except for 2014), as well as the changes in rates during the period. The 21 mutually exclusive causes of death were: 1) HIV/AIDS and tuberculosis, 2) diarrhea, lower respiratory and other common infectious diseases, 3) neglected tropical diseases and malaria, 4) maternal disorders, 5) neonatal disorders, 6) nutritional deficiencies, 7) other communicable, maternal, neonatal, and nutritional diseases (these seven causes are under the larger category of communicable, maternal, neonatal and nutritional diseases); 8) neoplasms, 9) cardiovascular diseases, 10) chronic respiratory diseases, 11) cirrhosis and other chronic liver diseases, 12) digestive diseases, 13) neurological disorders, 14) mental and substance use disorders, 15) diabetes, urogenital, blood and endocrine diseases, 16) musculoskeletal disorders, 17) other non-communicable diseases (these ten causes are under the larger category of non-communicable diseases); 18) transport injuries, 19) unintentional injuries, 20) self-harm and interpersonal violence, and 21) forces of nature, war and legal intervention (these four causes fall under the larger category of injuries) (Dwyer-Lindgren et al. 2016).

County-level information on population characteristics (including size, gender, age, and ethnicity), socio-economics (including educational level, median household income, unemployment and poverty rate), healthcare service (including medical insured rate and number of physicians per 1,000 population), and residential environment and location (including the Rural-Urban Continuum Code, and the latitude and longitude) were derived from the USA national official sources. The information on the studied counties was all complete.

In particular, population size, gender, age and ethnic distribution were collected from the US Bureau of Census (County Characteristics Resident Population Estimates, 2014 data). Education levels of adults were gathered from the US Bureau of Census (American Community Survey 5-year Average County-level Estimates, data of 2014-18). The data of annual median household income for each county was retrieved from the US Bureau of Census (Small Area Income and Poverty Estimates (SAIPE) Program, 2014 data). Unemployment rates for each county were collected from the US Bureau of Labor Statistics (Local Area Unemployment Statistics, 2014 data). Poverty rates were also collected from the US Census Bureau (SAIPE Program, 2014 data). The rates of insured population of medicine (aged under 65 years old) were derived from the US Census Bureau (Small Area Health Insurance Estimates Program, 2014 data). Number of physicians per 1,000 residents was collected from the database of Health Resources and Services Administration (Area Health Resources, 2014 data). The latest version of Rural-Urban Continuum Code, published in 2013 by the Economic Research Service of US Department of Agriculture, was used. Information on the location of each county was collected using coordinates (latitude and longitude) based on geographic centroids.

Databases of geochemistry, population health and other information collected were mapped together by county identity, where all variables were complete in 3080 counties (98.1% of the total counties). Counties with missing data included South Boston, Bedford City and Clifton Forge (in Virginia), Dade (in Florida), Shannon (in South Dakota), and a part of Yellowstone National Park (in Montana)). Data imputation for these counties was not attempted as the proportion of missing data was very small. Analyses followed the ‘complete-case’ approach, and thus 3080 county samples were included.

*Statistical analysis*

Health data in 2014 (i.e. life expectancy at birth, age-specific mortality risk, and cause-specific mortality rate; outcome variables) and geochemical data (i.e. concentrations of Al, As, Ca, Cu, Fe, P, Pb, Mg, Mn, Hg, Na, Se, Ti, and Zn; exposure variables) were analyzed. From the longitudinal perspective, the change in life expectancy between 1980 and 2014 was also used as an additional outcome variable, indicating the health improvement over time.

Descriptive and simple statistics (e.g. Spearman’s correlation test) were presented on the variables (including geochemical data, population characteristics, socio-economics, healthcare services, residential environment and location, and health statistics) were presented at first.

Univariate linear regression model was initially applied to show the association between each concentration of mineral and life expectancy at birth. Multivariate linear regression model with backward selection was then used to identify the final list of minerals associated with life expectancy at birth, with adjustment for other associated minerals.

Latent class analysis (LCA) was used to identify the common patterns of mineral distribution across the USA. LCA clustered counties were determined by the concentration features of the final listed minerals (derived from the previous backward selection process), and each county was allocated to one cluster. Before entering the data into LCA, the mineral concentration variables were first standardized and calculated using the Z-score method. The Log-likelihood (LL) statistics, Bayes Information Criterion (BIC), Consistent Akaike’s Information Criterion (CAIC), and the size of smallest cluster were applied to determine the optimal model (i.e. the optimal number of clusters). The software we used to process the analyses was Latent GOLD (version 4.5) with the estimation-maximization and Newton-Raphson algorithms applied to estimate model parameters. One thousand different random starting values which include 100 iterations for each run were used. Lower LL statistics suggests a better model fit. For the BIC and CAIC, the optimal model is defined as a model with the smallest information criterion values. However, to limit the total number of clusters and to ensure all clusters representing a significant proportion, we set that the size of smallest cluster should be 5% or more of the studied USA counties. Counties were allocated to clusters based on their posterior probabilities of belonging to each cluster. A mean posterior probability ≥ 0.7 for samples allocated to a cluster was considered acceptable (Alexander et al. 2020).

Multivariate linear regression analyses were conducted to examine the relationship between LCA derived clusters, and life expectancy (2014 data), change in life expectancy (between 1980 to 2014), age-specific mortality risk (2014 data) and cause-specific mortality rate (2014 data), with adjustment for population (except for age distribution), socio-economic, healthcare service, and residual environment and location factors. Age distribution was not included in adjustment since it is not naturally independent from the outcome variables measuring life span and death.

All the statistical analyses were carried out in the STATA 15 if not mentioned in a particular matter. Result visualization was done in R 4.1.2 and GraphPad Prism 9. A *p*-value < 0.005 (instead of 0.05), two tailed, was set as statistically significance, aiming to obtain conservative estimations of significant associations concerning the issue of multiple testing. A flowchart of the process of analysis is provided as the Supplementary Figure 1.

**Results**

Geochemical concentrations of all 14 minerals (Al, As, Ca, Cu, Fe, Hg, Mg, Mn, Na, P, Pb, Se, Ti and Zn), included in the National Geochemical Survey, were extracted from the original database, and are summarized for the 3080 studied counties in Table 1. Correlations of geochemical concentrations between these minerals are presented in Supplementary Table 1. The mean value of life expectancy at birth of the studied counties in 2014 was 77.75 years (standard deviation (SD), 2.37), and the mean value of increase in life expectancy between 1980 and 2014 was 5.34 years (SD, 1.99). Summarized statistics on age-specific mortality risk and cause-specific mortality rate are also shown in Table 1, as well as the socio-demographic information.

All 14 studied minerals were associated with life expectancy in univariate analyses (Table 2). However, after the process of backward selection in multivariate regression analyses, As, Ca, Na, Se and Zn remained their significant associations (Table 2).

LCA was carried out on these five minerals for common distribution patterns. The 3-cluster model was determined as the optimal model (Supplementary Table 2). Counties displayed very high posterior probabilities for their assigned clusters, with mean posterior probabilities ranging from 0.94 to 0.97 across the three clusters (Table 3). 66.8% (*n*=2056) of studied counties were assigned in a cluster where the concentrations of these five minerals were close to the overall mean (the ‘Common’ cluster, Figure 1a). The second cluster contained 24.0% (*n*=739) of studied counties, and was characterized by low concentrations of As, Ca, Na, Se and Zn (the ‘Infertile’ cluster, Figure 1b). Cluster 3 contained 9.3% (*n*=285) of studied counties, and in these regions As, Ca, Se and Zn concentrations were the highest (the ‘Plentiful’ cluster, Figure 1c).

In the southeast region of USA where most counties were assigned to the ‘Infertile’ cluster (Figure 2a), the life expectancy was the shortest (Figure 2b). No difference was seen in life expectancy between the ‘Common’ and ‘Plentiful’ clusters (Figure 2c). Life expectancies were consistently lower in the ‘Infertile’ cluster from 1980 to 2014, compared to the other two clusters (Figure 2c). It was 2.8 years lower based on the data of 2014 (unadjusted value). Furthermore, the increase in life expectancy from 1980 to 2014 was slower in the ‘Infertile’ cluster (Figure 2d), suggesting that health disparity was enlarging over time. County-level statistics of mineral concentrations, health measurements and socio-demographics, stratified by LCA derived clusters, is provided in Supplementary Table 3. After adjustment for the studied socio-demographic factors, the observed associations were still in highly significance, in models where the *R2* = 0.71 for life expectancy and 0.42 for increase in life expectancy, respectively (Table 4). Repeated analyses separately in counties classified as metro and non-metro areas (according to the Rural-Urban Continuum Code) demonstrated that such relationship was stronger in non-metro areas (Supplementary Tables 4 and 5).

Age-specific mortality risks, among all age categories, were the highest in the residents of ‘Infertile’ cluster. For example, for the group of 45-65 years the risks were 12.5% (SD 2.8%), 15.6% (2.7%) and 12.2% (2.7%) for those who lived in the counties of ‘Common’, ‘Infertile’ and ‘Plentiful’ clusters, respectively (Figure 3a and Supplementary Table 3). Adjusted difference in age-specific mortality risks between clusters was shown in Figure 3b (the ‘Common’ cluster as the referent group). As age increased, the proportion of additional mortality risk attributed to the ‘Infertile’ cluster was found to be decreasing (the ‘Common’ cluster as the referent group) (Figure 3c).

For cause-specific mortality, rates stratified by clusters suggested that the ‘Infertile’ cluster was associated with increased death from many causes (Supplementary Table 3). Adjusted results, where the ‘Common’ cluster was set as the referent, are shown in Figure 4. It was particularly worth noting that, compared to the ‘Common’ cluster, the rates of the top five most frequent causes of death (i.e. cardiovascular diseases, neoplasms, neurological disorders, chronic respiratory diseases, and diabetes, urogenital, blood and endocrine diseases) were all significantly higher in the ‘Infertile’ cluster (Figure 4). The differences in mortality rates from the 21 mutually exclusive causes among the three clusters over 1980 – 2014 are presented in Figure 5.

**Discussion**

Biological and medical research on the effects of minerals on certain organisms has made great progress (Lossow et al. 2021) (Huybrechts et al. 2021), although it is not yet clear which minerals are essential or beneficial, and which are non-essential or even toxic, particularly for humans (Maret 2016). In the field of environmental research, geographically large-scale studies on the spatial distribution of minerals in samples such as soil and water are common; however, there is a lack of corresponding assessments of health effects on human populations (Hu et al. 2020). On the other hand, epidemiological studies of human populations have typically been conducted only at a small region on specific minerals and diseases (J. Wang et al. 2021), due to difficulties in study design, cost, management and mapping of mineral exposures and population health outcomes. Due to the open-access, high-quality databases (Smith et al. 2013) (Dwyer-Lindgren et al. 2017) (Dwyer-Lindgren et al. 2016), we have been able to conduct this study on a national scale in the USA, taking into account some new aspects which filled some important research gaps.

Concentrations of As, Ca, Na, Se and Zn in surface stream-sediment or soil were found to be associated with life expectancy at birth in a large geographical matter. Ca, Na, Se and Zn were in positive associations, whereas As was negative. For the other studied minerals (Al, Cu, Fe, Hg, Mg, Mn, P, Pb and Ti), our multivariate regression models suggested that those apparent associations initially occurred in simple statistics could be due to their co-existence with Ca, Na, Se or Zn, suggesting to the need to carry out investigations including multiple minerals simultaneously.

For the first time, using LCA, the common patterns of geographical distribution of As, Ca, Na, Se and Zn across USA was identified where three distinctive clusters were determined. Previous studies have described the spatial distributions of specific minerals separately (Hu et al. 2020) (Woodruff et al. 2015) (Rembert et al. 2017), but to our knowledge no one has considered the patterns of multiple minerals in an integrated way, particularly with regard to human health.

The “Infertile” cluster, which were counties characterized by the lowest concentrations of As, Ca, Na, Se and Zn, was associated with a reduced life expectancy of their residents by approximately 3 years (2014, unadjusted data). With adjustment for other collected variables, including population characteristics, socio-economics, healthcare service, and residual environment and location, this relationship remained highly significant. This cluster was also associated with the smallest increase in life expectancy between 1980 and 2014 compared to the remaining clusters, further supporting its vulnerability from the view of longitudinal perspective. These results are important as this cluster accounts for approximately one quarter of the total numbers of USA counties. Particularly, geographically this cluster largely overlaps the areas where life expectancy of their residents was the shortest (Figure 2a and b). No difference in life expectancy or increase in life expectancy between 1980 and 2014 was found between the ‘Common’ and ‘Plentiful’ clusters, implying a sufficient amount of essential minerals rather than a higher amount is the key to maintain good health.

Through further analysis of age-specific mortality risk, we found that counties labeled as the 'Infertile' cluster were associated with higher mortality risks at all age groups. Particularly, the risks of mortality attributed to the 'Infertile' cluster were proportionally lager in populations who were younger. These results suggest that any insufficiency of Ca, Na, Se or Zn in the environment may have negative impacts on the health of local population throughout the lifetime.

We studied the mortality rates for 21 mutually exclusive causes of death, and found that the 'Infertile' cluster was associated with higher mortality rates for many causes. In particular, increased rates were observed for the top five leading causes of death (i.e. cardiovascular diseases, neoplasms, neurological disorders, chronic respiratory diseases, and diabetes, urogenital, blood and endocrine diseases). Previous studies provided varying degrees of evidence for the effects of these individual elements on specific diseases or health outcomes (Zoroddu et al. 2019) (Dinh et al. 2018) (Hasan 2021) (Eaton et al. 2010) (Benstoem et al. 2015) (Little et al. 2010) (Quansah et al. 2015) (A. Rahman et al. 2010) (Jackson 1988) (Cavdar et al. 2009) (Kazi et al. 2012) (Guo et al. 2012) (Y. X. Wang et al. 1990) (Zaichick VYe et al. 1995) (Adaramoye et al. 2010) (Ma et al. 2020), for example, the association of Zn and Se deficiency with cardiovascular diseases such as myocarditis, arrhythmias and coronary artery disease (Eaton et al. 2010) (Benstoem et al. 2015) (Powell et al. 1995) (Coudray et al. 1993), and with neoplasms such as stomach cancer, liver cancer, bladder cancer, prostate cancer and malignant lymphomas (Jackson 1988) (Cavdar et al. 2009) (Kazi et al. 2012) (Guo et al. 2012) (Y. X. Wang et al. 1990) (Adaramoye et al. 2010). However, there has been no previous research looking at the patterns of multiple minerals in the environment in a geographical region, and investigated their collective effects on different diseases and conditions. Our study, as the first of this type, provides novel insights on the spatial relationship between environmental minerals and human health.

In the literature, many studies have demonstrated the substantial and still increasing geographical disparities in life expectancy in the USA (Dwyer-Lindgren et al. 2017) (Dwyer-Lindgren et al. 2016) (Wei et al. 2012) (Kulkarni et al. 2011) (Chang et al. 2015) (Boing et al. 2020). Based on the data of 1999 – 2001, Mississippi was associated with the shortest life expectancy at birth (73.9 years) and Hawaii the longest (80.2) (Wei et al. 2012). Relative to a larger area, populations in the South region were associated with the shortest life expectancy, while in the Northeast the longest (Chang et al. 2015). Health disparities between regions continued to widen: in 2007 the difference between counties with the highest and lowest life expectancy at birth was 12.5 years for females and 15.2 years for males (Kulkarni et al. 2011), whereas in 2014 the difference enlarged to 20.1 years (summarized for both genders) (Dwyer-Lindgren et al. 2017). Difference in life expectancy should refer to disparities in the occurrence and outcome of diseases. Analyses of cause-specific mortality in the USA showed that such disparities by geographical location were very large. For example, for cardiovascular diseases, the top cause of death in the USA, it showed a seven-fold difference in mortality risks between the highest and lowest counties (503.1 vs. 70.7 per 100,000 population, data of 2014) (Dwyer-Lindgren et al. 2016). Regional factors associated with health disparities were mainly analyzed and discussed in the domains of population demographics, socio-economics and healthcare service (Dwyer-Lindgren et al. 2017) (Dwyer-Lindgren et al. 2016) (Chang et al. 2015) (Boing et al. 2020) (Vierboom et al. 2019) (Arora et al. 2016). Our study, at the USA national scale, provides evidence, in addition to the above domains, that geochemistry may also be a very important group of factors. In future studies, investigations on the influence of prenatal and early exposures to infertile environments on later health and development are of great interest.

The availability of mineral and health data across the USA offers a novel research opportunity (i.e. linked database research) to investigate environmental epidemiological issues and explore the relationship between environmental exposure and human health, at a very large geographical scale. An ecological study, using data reported at the county level, is a feasible and effortful approach, since that in the USA counties are the smallest administrative units where essential studied information can be provided (Dwyer-Lindgren et al. 2017). This study benefits from the fact that the USA is a well-developed country, across which information on population demographics, socio-economics, healthcare service and statistics of residual environment is well recorded by the Federal offices. It enabled us to carry out a relatively comprehensive adjustment for potential confounding factors in our analyses. It is worth noting that our studied variables together, based on our regression model, can explain more than 70% of the variance of county-level statistics of life expectancy (*R2* = 0.711, Table 4). However, many other variables, such as food and drinking water sources, are not available, which is an obvious limitation of this study.

We were aware that the food and drinking water consumed may differ in rural *versus* urban areas, with the former being more likely to consume locally produced products as compared to the latter where more globalized and processed products are used. Thus, the Rural-Urban Continuum Code, which served as a surrogate for the difference between rural and urban areas, was included for statistical adjustment in our analyses. Furthermore, we conducted separate analyses by metro and non-metro areas, and found that the association between mineral cluster and life expectancy was weaker in metro areas (Supplementary Tables 4 and 5).

Mean concentrations of minerals at the county level are the only data accessible from the National Geochemical Survey. Additional information, such as other parameters of samples that determine the bioavailability of the analyzed minerals, is not available. The geochemical database currently includes only14 minerals, and thus our results are based on this list. Other potentially important minerals such as Co, I, K and Mo were not analyzed.

Indicators of overall population health such as life expectancy and mortality, were used in this study, instead of more specific measurements such as the status of certain diseases. It should be noted that there may be [non-negligible](https://www.powerthesaurus.org/non-negligible/synonyms) variations among the standards of disease identification and recording across the USA, and unbiased information of specific health measurements is scarce. Furthermore, many health indicators would be affected by detection capability and suffered from survival bias. For example, higher prevalence of lung cancer in a region could actually be attributed to a better method of cancer diagnosis, or a better treatment leading to a longer survival period. However, life expectancy and mortality data (from the national death registration database with information on the cause of death) are less problematic particularly in geographically large-scale research across multiple administration areas, and they provide a less unbiased evaluation on overall population health.

As an ecological study, common issues, such as migration bias, may have their impacts. Our time series analysis from 1980 to 2014 (Figure 5), however, suggested that the findings were largely consistent over the decades, ruling out a strong effect. We are unable to rule out potential reporting bias which may exist across different counties, while the data used in this study are considered from high quality sources (Smith et al. 2013) (Dwyer-Lindgren et al. 2017) (Dwyer-Lindgren et al. 2016). Nevertheless, particular attention must be paid to ecological fallacy and therefore no inferences are intended to be made at the individual level (Sedgwick 2014). The geographical pattern of minerals in this study is determined by LCA with the mentioned selection criteria for the optimal model, however, use of other analytical approaches may generate slightly different results.

**Conclusions**

Our county-level ecological study identified the geographical patterns of life expectancy-related minerals (As, Ca, Na, Se and Zn) in stream-sediment (or surface soil) samples across the contiguous USA, where three distinctive clusters were determined. In particular, the cluster, characterized by the regions having the lowest concentrations of these five elements, was associated with a lower life expectancy at birth, higher mortality risks at all ages, and higher mortality rates of many specific reasons including the five leading causes of death. Our study brings new perspectives to explain health disparities in the USA.

**Statements & Declarations**

Funding: This study was supported by the Postgraduate Research Studentship (PGRS, no. 2012021) Award at Xi’an Jiaotong-Liverpool University.

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions: YC, LY, RG and ZFM conceived the study. BQ, SW, YC and PZ analyzed the data. YC drafted the initial manuscript. All authors contributed to the study design and interpretation of the data, and all authors approved the final version of the manuscript submitted for publication. BQ and SW equally contributed to this paper.

Ethics approval**:** Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

**Bibliography**

Adaramoye, O. A., Akinloye, O., & Olatunji, I. K. (2010). Trace elements and vitamin E status in Nigerian patients with prostate cancer. *African Health Sciences*, *10*(1), 2–8.

Alexander, J., Tinkov, A., Strand, T. A., Alehagen, U., Skalny, A., & Aaseth, J. (2020). Early Nutritional Interventions with Zinc, Selenium and Vitamin D for Raising Anti-Viral Resistance Against Progressive COVID-19. *Nutrients*, *12*(8), 2358. https://doi.org/10.3390/nu12082358

Alloway, B. J. (2009). Soil factors associated with zinc deficiency in crops and humans. *Environmental Geochemistry and Health*, *31*(5), 537–548. https://doi.org/10.1007/s10653-009-9255-4

Arora, A., Spatz, E., Herrin, J., Riley, C., Roy, B., Kell, K., et al. (2016). Population Well-Being Measures Help Explain Geographic Disparities In Life Expectancy At The County Level. *Health Affairs (Project Hope)*, *35*(11), 2075–2082. https://doi.org/10.1377/hlthaff.2016.0715

Benstoem, C., Goetzenich, A., Kraemer, S., Borosch, S., Manzanares, W., Hardy, G., & Stoppe, C. (2015). Selenium and Its Supplementation in Cardiovascular Disease—What do We Know? *Nutrients*, *7*(5), 3094–3118. https://doi.org/10.3390/nu7053094

Boing, A. F., Boing, A. C., Cordes, J., Kim, R., & Subramanian, S. V. (2020). Quantifying and explaining variation in life expectancy at census tract, county, and state levels in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(30), 17688–17694. https://doi.org/10.1073/pnas.2003719117

Cavdar, A. O., Gözdaşoğlu, S., Babacan, E., Mengübaş, K., Unal, E., Yavuz, G., & Taçyildiz, N. (2009). Zinc and selenium status in pediatric malignant lymphomas. *Nutrition and Cancer*, *61*(6), 888–890. https://doi.org/10.1080/01635580903285189

Chang, M.-H., Molla, M. T., Truman, B. I., Athar, H., Moonesinghe, R., & Yoon, P. W. (2015). Differences in healthy life expectancy for the US population by sex, race/ethnicity and geographic region: 2008. *Journal of Public Health (Oxford, England)*, *37*(3), 470–479. https://doi.org/10.1093/pubmed/fdu059

Coudray, C., Charlon, V., de Leiris, J., & Favier, A. (1993). Effect of zinc deficiency on lipid peroxidation status and infarct size in rat hearts. *International Journal of Cardiology*, *41*(2), 109–113. https://doi.org/10.1016/0167-5273(93)90149-B

Dinh, Q. T., Cui, Z., Huang, J., Tran, T. A. T., Wang, D., Yang, W., et al. (2018). Selenium distribution in the Chinese environment and its relationship with human health: A review. *Environment International*, *112*, 294–309. https://doi.org/10.1016/j.envint.2017.12.035

Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Kutz, M. J., Huynh, C., et al. (2016). US County-Level Trends in Mortality Rates for Major Causes of Death, 1980-2014. *JAMA*, *316*(22), 2385–2401. https://doi.org/10.1001/jama.2016.13645

Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Mackenbach, J. P., van Lenthe, F. J., et al. (2017). Inequalities in Life Expectancy Among US Counties, 1980 to 2014: Temporal Trends and Key Drivers. *JAMA internal medicine*, *177*(7), 1003–1011. https://doi.org/10.1001/jamainternmed.2017.0918

Eaton, C. B., Abdul Baki, A. R., Waring, M. E., Roberts, M. B., & Lu, B. (2010). The association of low selenium and renal insufficiency with coronary heart disease and all-cause mortality: NHANES III follow-up study. *Atherosclerosis*, *212*(2), 689–694. https://doi.org/10.1016/j.atherosclerosis.2010.07.008

Farag, M. A., Hamouda, S., Gomaa, S., Agboluaje, A. A., Hariri, M. L. M., & Yousof, S. M. (2021). Dietary Micronutrients from Zygote to Senility: Updated Review of Minerals’ Role and Orchestration in Human Nutrition throughout Life Cycle with Sex Differences. *Nutrients*, *13*(11), 3740. https://doi.org/10.3390/nu13113740

Gashu, D., Nalivata, P. C., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., et al. (2021). The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature*. https://doi.org/10.1038/s41586-021-03559-3

Guo, K.-F., Zhang, Z., Wang, J.-Y., Gao, S.-L., Liu, J., Zhan, B., et al. (2012). Variation of urinary and serum trace elements (Ca, Zn, Cu, Se) in bladder carcinoma in China. *Asian Pacific journal of cancer prevention: APJCP*, *13*(5), 2057–2061. https://doi.org/10.7314/apjcp.2012.13.5.2057

Hambidge, M. (2000). Human zinc deficiency. *The Journal of Nutrition*, *130*(5S Suppl), 1344S–9S. https://doi.org/10.1093/jn/130.5.1344S

Hasan, S. E. (2021). Medical Geology. *Encyclopedia of Geology*, 684–702. https://doi.org/10.1016/B978-0-12-409548-9.12523-0

Hu, B., Shao, S., Ni, H., Fu, Z., Hu, L., Zhou, Y., et al. (2020). Current status, spatial features, health risks, and potential driving factors of soil heavy metal pollution in China at province level. *Environmental Pollution (Barking, Essex: 1987)*, *266*(Pt 3), 114961. https://doi.org/10.1016/j.envpol.2020.114961

Huybrechts, M., Hendrix, S., Kyndt, T., Demeestere, K., Vandamme, D., & Cuypers, A. (2021). Short-term effects of cadmium on leaf growth and nutrient transport in rice plants. *Plant Science: An International Journal of Experimental Plant Biology*, *313*, 111054. https://doi.org/10.1016/j.plantsci.2021.111054

Jackson, M. L. (1988). Selenium: geochemical distribution and associations with human heart and cancer death rates and longevity in China and the United States. *Biological Trace Element Research*, *15*, 13–21. https://doi.org/10.1007/BF02990124

Kazi, T. G., Kolachi, N. F., Afridi, H. I., Kazi, N. G., Sirajuddin, null, Naeemullah, null, & Arain, S. S. (2012). Effects of mineral supplementation on liver cirrhotic/cancer male patients. *Biological Trace Element Research*, *150*(1–3), 81–90. https://doi.org/10.1007/s12011-012-9501-y

Kulkarni, S. C., Levin-Rector, A., Ezzati, M., & Murray, C. J. (2011). Falling behind: life expectancy in US counties from 2000 to 2007 in an international context. *Population Health Metrics*, *9*(1), 16. https://doi.org/10.1186/1478-7954-9-16

Little, P. J., Bhattacharya, R., Moreyra, A. E., & Korichneva, I. L. (2010). Zinc and cardiovascular disease. *Nutrition*, *26*(11), 1050–1057. https://doi.org/10.1016/j.nut.2010.03.007

Lossow, K., Renko, K., Schwarz, M., Schomburg, L., Schwerdtle, T., & Kipp, A. P. (2021). The Nutritional Supply of Iodine and Selenium Affects Thyroid Hormone Axis Related Endpoints in Mice. *Nutrients*, *13*(11), 3773. https://doi.org/10.3390/nu13113773

Ma, J., Yan, L., Guo, T., Yang, S., Liu, Y., Xie, Q., et al. (2020). Association between Serum Essential Metal Elements and the Risk of Schizophrenia in China. *Scientific Reports*, *10*(1), 10875. https://doi.org/10.1038/s41598-020-66496-7

Maret, W. (2016). The Metals in the Biological Periodic System of the Elements: Concepts and Conjectures. *International Journal of Molecular Sciences*, *17*(1), E66. https://doi.org/10.3390/ijms17010066

Medical Geology and the Soil Health-Human Health Nexus. (2019). *Environmental and Pollution Science*, 501–510. https://doi.org/10.1016/B978-0-12-814719-1.00027-6

Park, S. K., Sack, C., Sirén, M. J., & Hu, H. (2020). Environmental Cadmium and Mortality from Influenza and Pneumonia in U.S. Adults. *Environmental Health Perspectives*, *128*(12), 127004. https://doi.org/10.1289/EHP7598

Powell, S. R., Aiuto, L., Hall, D., & Tortolani, A. J. (1995). Zinc supplementation enhances the effectiveness of St. Thomas’ Hospital no. 2 cardioplegic solution in an in vitro model of hypothermic cardiac arrest. *The Journal of Thoracic and Cardiovascular Surgery*, *110*(6), 1642–1648. https://doi.org/10.1016/S0022-5223(95)70025-0

Quansah, R., Armah, F. A., Essumang, D. K., Luginaah, I., Clarke, E., Marfoh, K., et al. (2015). Association of arsenic with adverse pregnancy outcomes/infant mortality: a systematic review and meta-analysis. *Environmental Health Perspectives*, *123*(5), 412–421. https://doi.org/10.1289/ehp.1307894

Rahman, A., Persson, L.-Å., Nermell, B., El Arifeen, S., Ekström, E.-C., Smith, A. H., & Vahter, M. (2010). Arsenic exposure and risk of spontaneous abortion, stillbirth, and infant mortality. *Epidemiology (Cambridge, Mass.)*, *21*(6), 797–804. https://doi.org/10.1097/EDE.0b013e3181f56a0d

Rahman, S. M., Kippler, M., Tofail, F., Bölte, S., Hamadani, J. D., & Vahter, M. (2017). Manganese in Drinking Water and Cognitive Abilities and Behavior at 10 Years of Age: A Prospective Cohort Study. *Environmental Health Perspectives*, *125*(5), 057003. https://doi.org/10.1289/EHP631

Rembert, N., He, K., Judd, S. E., & McClure, L. A. (2017). The geographic distribution of trace elements in the environment: the REGARDS study. *Environmental Monitoring and Assessment*, *189*(2), 84. https://doi.org/10.1007/s10661-016-5733-1

Schullehner, J., Thygesen, M., Kristiansen, S. M., Hansen, B., Pedersen, C. B., & Dalsgaard, S. (2020). Exposure to Manganese in Drinking Water during Childhood and Association with Attention-Deficit Hyperactivity Disorder: A Nationwide Cohort Study. *Environmental Health Perspectives*, *128*(9), 97004. https://doi.org/10.1289/EHP6391

Sedgwick, P. (2014). Ecological studies: advantages and disadvantages. *BMJ (Clinical research ed.)*, *348*, g2979. https://doi.org/10.1136/bmj.g2979

Smith, D. B., Smith, S. M., & Horton, J. D. (2013). History and evaluation of national-scale geochemical data sets for the United States. *Geoscience Frontiers*, *4*(2), 167–183. https://doi.org/10.1016/j.gsf.2012.07.002

Vierboom, Y. C., Preston, S. H., & Hendi, A. S. (2019). Rising geographic inequality in mortality in the United States. *SSM - population health*, *9*, 100478. https://doi.org/10.1016/j.ssmph.2019.100478

Wang, J., Wang, X., Li, H., Yang, L., Li, Y., & Kong, C. (2021). Spatial distribution and determinants of health loss from Kashin-Beck disease in Bin County, Shaanxi Province, China. *BMC public health*, *21*(1), 387. https://doi.org/10.1186/s12889-021-10407-6

Wang, Y. X., Qin, J. F., Wu, S. M., & Yan, L. B. (1990). Study on the relation of Se, Mn, Fe, Sr, Pb, Zn, Cu, and Ca to liver cancer mortality from analysis of scalp hair. *The Science of the Total Environment*, *91*, 191–198. https://doi.org/10.1016/0048-9697(90)90298-9

Wei, R., Anderson, R. N., Curtin, L. R., & Arias, E. (2012). U.S. decennial life tables for 1999-2001: state life tables. *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, *60*(9), 1–66.

Woodruff, L., Cannon, W. F., Smith, D. B., & Solano, F. (2015). The distribution of selected elements and minerals in soil of the conterminous United States. *Journal of Geochemical Exploration*, *154*, 49–60. https://doi.org/10.1016/j.gexplo.2015.01.006

Zaichick VYe, null, Tsyb, A. F., & Vtyurin, B. M. (1995). Trace elements and thyroid cancer. *The Analyst*, *120*(3), 817–821. https://doi.org/10.1039/an9952000817

Zha, X., An, J., Gao, X., & Tian, Y. (2022). Dietary and drinking water intake of essential trace elements in a typical Kashin-Beck disease endemic area of Tibet, China. *Environmental Health: A Global Access Science Source*, *21*(1), 86. https://doi.org/10.1186/s12940-022-00898-0

Zhang, X., Wang, T., Li, S., Ye, C., Hou, J., Li, Q., et al. (2019). A Spatial Ecology Study of Keshan Disease and Hair Selenium. *Biological Trace Element Research*, *189*(2), 370–378. https://doi.org/10.1007/s12011-018-1495-7

Zoroddu, M. A., Aaseth, J., Crisponi, G., Medici, S., Peana, M., & Nurchi, V. M. (2019). The essential metals for humans: a brief overview. *Journal of Inorganic Biochemistry*, *195*, 120–129. https://doi.org/10.1016/j.jinorgbio.2019.03.013