

Spatial clusters of life expectancy and association with cardiovascular disease mortality and cancer mortality in the contiguous United States: 1980-2014

Raid W. Amin, Julia Steinmetz

Department of Mathematics and Statistics, University of West Florida, Pensacola, FL, USA

Abstract

The average life expectancy varies greatly from county to county in USA and there are also spatial variations in the county mortality rates for cardiovascular disease (CVD) and cancer, the top two causes of death. An association between these two groups of diseases has not been identified by cluster analysis previously. The main objective in this study was to investigate and quantify the associations between mortality due to CVD, cancer mortality and life expectancy based on US county data between 1980 and 2014. Regression analysis was used to adjust life expectancy for the mortality due to CVD and that due to cancer. In addition to the spatial life expectancy trends, we also studied existing trends over time with the software JOINPOINT to see how life expectancy is influenced by changes in mortality due to CVD and cancer mortality. The study setting was the 48 contiguous US states, while participants were 3,100 counties and their populations of all ages during the period 1980-2014. The main outcomes are spatial clusters of unusually low or high levels of life expectancy in addition to identifying which county level life expectancy locations were

significantly associated with mortality due to CVD and/or cancer. Life expectancy has been improving steadily from 1980 to 2014, but the rate of increase per year (indicated by variation of the trend slope) changed significantly at five joinpoints, the latest of which occurred in 2010 when the slope changed from 0.29 (1980-1982) to 0.03 (2010-2014). Our results reveal that there are significant, purely spatial clusters in some geographical areas where life expectancy rates are significantly higher (or lower) than in the rest of the contiguous US. It is also shown that there is a significant association between the life expectancy level and the corresponding CVD mortality, and there is also a significant association between life expectancy level and the corresponding overall cancer mortality. The general trends (regression slopes) over time for the USA in life expectancy mortality, CVD mortality and cancer mortality have changed significantly after 2009-2010.

Introduction

Significant inequalities with respect to cardiovascular disease (CVD) exist between the counties in USA (Dwyer-Lindgren *et al.*, 2016, 2017), and there is an association between income and life expectancy (Tarkiainen *et al.*, 2012; Chetty *et al.*, 2016). A significant association has been found between concentration of fine particulate matter and life expectancy (Pope *et al.*, 2009, Wang *et al.*, 2014). It has also been shown that inequality in income is associated with mortality in the USA (Kaplan *et al.*, 1996), and that life expectancy is associated with the level of education attained (Meara *et al.*, 2008). Race also plays a role in life expectancy (Deaton and Lubotsky, 2003; Olshansky *et al.*, 2012). Therefore, life expectancy depends on several factors and it might even be difficult to find factors that do not have such effects. However, a national population-based study of the counties in USA on life expectancy and mortality due to heart disease, stroke, chronic obstructive pulmonary disease and various cancers concluded that living at high altitudes is one of the few variables without impact (Ezzati *et al.*, 2012).

This project focuses on the two main causes of death in USA, *i.e.* CVD and cancer. The Institute for Health Metrics and Evaluation (IHME), an independent research institute at the University of Washington working in the area of global health statistics and impact evaluation, publishes numerous publications about life expectancy and mortality (Dwyer-Lindgren *et al.*, 2016, 2017; Mokdad *et al.*, 2017). However, there are no published cluster analyses on the interrelationship between life expectancy and the mortality rates due to cancer and CVD. We were interested in identifying significant statistical associations, including purely spatial clusters, based on life expectancy and mortality due to these two disease groups based on county-level data for the 48

Correspondence: Raid W. Amin, Department of Mathematics and Statistics, University of West Florida, Building 4, Room 336, Pensacola, FL 32514, USA.
Tel.: 850.474.3014.
E-mail: ramin@uwf.edu

Key words: Life expectancy; Cluster analysis; Geospatial mapping; Cardiovascular disease; Cancer; USA.

Contributions: RA oversaw and approved the statistical modelling with SaTScan and ArcMap in addition to choosing; approved and interpreted all data analysis; contributed and approved the manuscript write-up. JS applied SaTScan and ArcMap, in addition to contributing to the write-up of the manuscript.

Conflict of interest: the authors declare no potential conflict of interest.

Funding: none.

©Copyright R.W. Amin and J. Steinmetz, 2019
Licensee PAGEPress, Italy
Geospatial Health 2019; 14:733
doi:10.4081/gh.2019.733

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

contiguous states in USA. The specific main research questions in this study were the following: i) are there geographical areas where life expectancy rates are significantly higher (or lower) than in the rest of the contiguous USA? Are there significant purely spatial clusters? ii) is there an association between the level of life expectancy and that of CVD mortality? Can this be identified spatially? iii) is there an association between the level of life expectancy and that of overall cancer mortality? Can this be identified spatially? iv) have the trends (regression slopes) with respect to life expectancy, CVD mortality and cancer mortality significantly changed over the period 1980-2014?

Materials and Methods

Life expectancy data, mortality data for cancer and age-adjusted CVD mortality data were obtained from the IHME, which used a small area estimation methodology to calculate the estimated mortality data for the years 1980-2014. It is the most comprehensive dataset available for mortality in the USA, and we therefore used the data for this whole period.

Since one of our goals was to find existing associations between life expectancy and mortality due to CVD and cancer in the counties, linear regression was used to adjust the life expectancy data for CVD mortality and cancer mortality. The latter included 29 different types of cancer, such as cancer of the brain and nervous system, thyroid, larynx, nasopharynx, other pharynx, trachea, bronchus and lung, lip and oral cavity, oesophagus, stomach, liver, gallbladder and biliary tract, pancreas, colon and rectum, kidney, bladder, skin (except melanoma), breast, ovary, cervix uterus, prostate, testis as well as Hodgkin and non-Hodgkin lymphoma, leukaemia, malignant skin melanoma, mesothelioma, multiple myeloma, other neoplasms.

We were also interested in the trends of life expectancy, mortality and incidence of disease. To that end we downloaded from the National Cancer Institute the software JOINPOINT (Kim *et al.*, 2000). JOINPOINT uses a nonparametric permutation based test for each time point to test the hypothesis that the slope has not changed versus that there exists a true change in the slope for the linear model. If the test indicates a change in the slope, then we have identified a joinpoint at a given time point at which the slope changes and at which a model then can obtain some other fit that is better than the one that was used until the joinpoint. Sometimes, there is no joinpoint identified. The permutation method randomly permutes the straight-line regression model based residuals, meaning that we shuffle around the distances between the regression line and each observation to calculate the test statistic. The sequence of permutation tests is used to ensure that the approximate probability of overall Type I error is less than the specified significance level. The overall Type I error is the probability of incorrectly concluding that the underlying regression model has at least one joinpoint when the true underlying model has no joinpoints. The test statistic measures the observed evidence that the data provides against the null hypothesis, and we obtain p-values on each tested time point. If the tests are significant, then there exists at least one joinpoint at which the true slope has changed. The number of joinpoints can range from none to several within a single regression.

Several permutation tests are performed to test for any significant changes in the trend slope(s) over time. The P-values for these tests are found with a Monte Carlo method (Kim *et al.*, 2000), and

the significance levels stay correct through a Bonferroni correction (Kim *et al.*, 2000).

Turning to cluster analysis, the disease surveillance software SaTScan™ (Kulldorff, 1997; Kulldorff and Information Management Services Inc., 2009) was used to identify purely, spatial clusters using a weighted normal model (Kulldorff *et al.*, 2009). We tested the null hypothesis that the age-adjusted rates of life expectancy, mortality due to CVD and cancer mortality are the same for all counties in contiguous USA. The spatial scan statistic in SaTScan™ identifies clusters by imposing a *scan window* that moves over a map including different sets of neighbouring counties represented by their corresponding geographic centroids. The purely spatial scan statistic places a circular window on the map, where the window is centered on each of several possible grid points positioned throughout the study region. The radius of the window for each grid point varies continuously in size, where the circular windows are flexible both in location and size. This scan method creates an infinite number of distinct geographical circles with different sets of neighbouring data locations within them. Each circle is a possible candidate cluster. The test is done by gradually scanning a window across time and/or space, noting the number of observed and expected observations inside the window at each location. If the window includes the centroid of a specific county, then this county will be accepted as included in the window. The centre of the window is positioned only at the county centroids. For each window, the spatial scan statistic tests the null hypothesis of equal rates against the alternative hypothesis that one or several of the parameters (here life expectancy, mortality due to CVD and cancer mortality) have different values within the scan window compared to areas outside the window.

Regression analysis was used to adjust mortality rates for CVD and for cancer. Since the mortality rates are continuous, we predicted these rates with a linear regression model from the covariates (or risk factors) CVD and cancer, and we used the resulting residuals as the *adjusted mortality rates*. These residuals were normalized and then analysed with SaTScan for a purely spatial cluster analysis. Changes between age adjusted mortality rates and age plus covariate adjusted mortality rates could then be identified in the cluster maps.

Results

Figures 1-3 show the different joinpoints for the three entries studied over the study period (1980-2014). Figure 1 shows 5 significant joinpoints identified by the JOINPOINT algorithm for life expectancy (both genders). However, the regression slope experiences only minor changes of direction, except at the last joinpoint in 2010 when the slope changes from a general upward trajectory with a slope of 0.29 for 1980-1982 to 0.03 for the remaining study years (2010-2014). This means that life expectancy used to increase each year by around 0.29 for the years up to 1982, and then dropped to an increase of only 0.03 years for the remaining part of the study years (2010-2014).

Figure 2 shows the regression slope with 4 joinpoints for CVD mortality, which seems to be decreasing steadily from 1980-2009, but then changes abruptly from a slope value of -9.61 to one of -0.92 for 2009-2014. With respect to cancer (Figure 3), there seems to have been a slight increase in mortality until 1995 (first joinpoint) after which mortality decreases strongly until the last joinpoint in 2010 where the slope changes from -2.80 to 0.52. In each



Figure 1. JOINPOINT analysis of life expectancy in the contiguous United States 1980-2014.

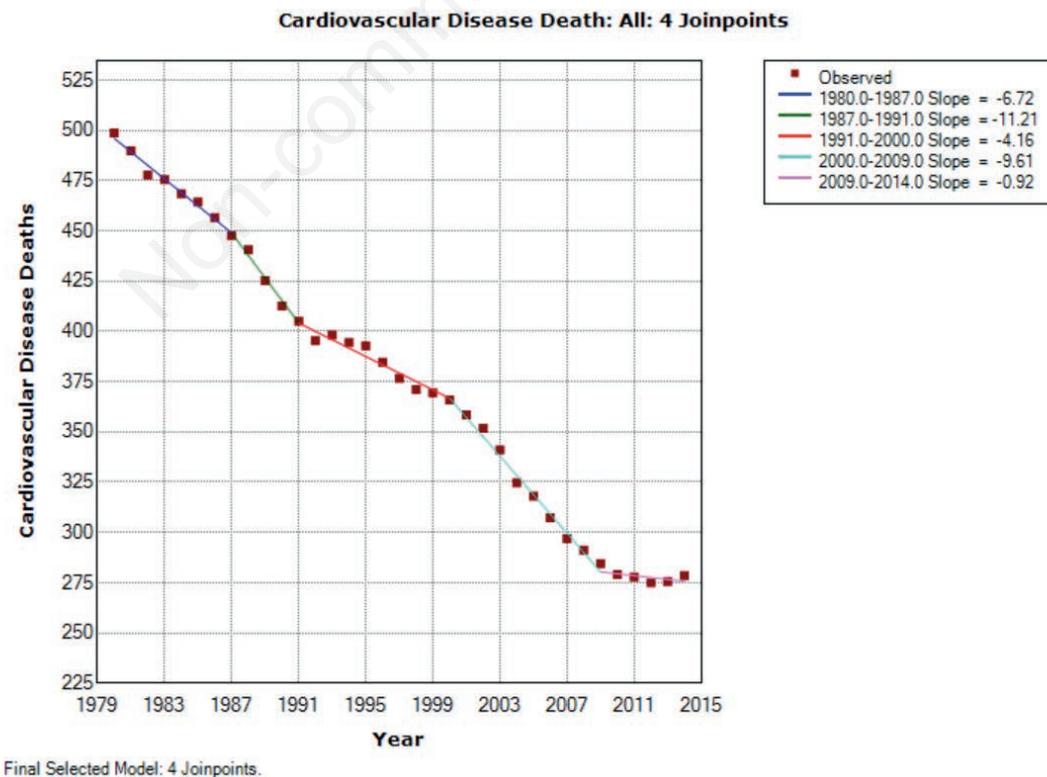


Figure 2. JOINPOINT analysis of cardiovascular disease mortality in the contiguous United States per 100,000 population over the period 1980-2014.

of these three joinpoint graphs the slope changes to a value close to zero in the very short period 2009-2010.

The heat map in Figure 4 gives each county colour shades according to the level of life expectancy, where higher values are represented by darker shades of blue (mainly in the Northwest) and lower values by darker shades of red (mainly in the Southeast). The circular shaped shaded areas on this map identified significant clusters of high levels of life expectancy (blue), mainly in the West, and clusters of low such levels (red) in the Southeast. While the *red* cluster in the Southeast includes counties with average life expectancy ranging from 64-73 years, the *blue* clusters in some parts of the USA include counties with average life expectancy ranging from 73-83 years. The large *blue* cluster in the West represents a mean life expectancy that is 0.63 standard errors greater than what was found in the rest of the contiguous USA, while the *red* cluster in the SE of the US has a mean life expectancy that is 1.21 standard errors lower than what is found in the rest of the contiguous US. These are the only two very large-size, significant clusters for life expectancy. There exist also very small clusters that are based on very few counties.

In order to capture any significant associations with cancer mortality on the life expectancy map, Figure 5 shows the identification of clusters of life expectancy after adjusting these rates for the overall cancer mortality rates. Orange refers to counties in which a low life expectancy is not associated with cancer mortality. Other factors may be affecting life expectancy there. Similarly, blue stands for those counties in which high life expectancy is not

associated with cancer mortality. Counties depicted in yellow emerged as low-level life expectancy clusters only after adjustment had been done for cancer mortality. The *dark green* counties are those with high life expectancy before we adjusted for cancer. The *blue* cluster shown in Figure 4 is now shown as dark green after we adjusted for cancer rates. The *light green* clusters seen in Figure 5 are counties with high life expectancy only after we had adjusted it for cancer rates. They were not inside any cluster until we adjusted life expectancy for cancer. Counties characterized as having a high life expectancy, but which turned into counties with a low life expectancy after cancer mortality had been taken into account. The counties in pink are those in which the low life expectancy was found to be associated with high cancer mortality. Figure 6 shows life expectancy clusters after adjusting for CVD mortality. The colours were assigned in the same way as used for Figure 5. The counties in pink here are those found to have a low life expectancy associated with high rates of CVD mortality. This is an effective method for clarifying the covariate adjustment effect on life expectancy on the same cluster maps.

Discussion and Conclusions

As identification of significant changes in mortality of a specific disease is an important issue, we paid particular attention to such changes. The negative correlation coefficient (r) between life expectancy and cancer mortality in the US has been strengthening

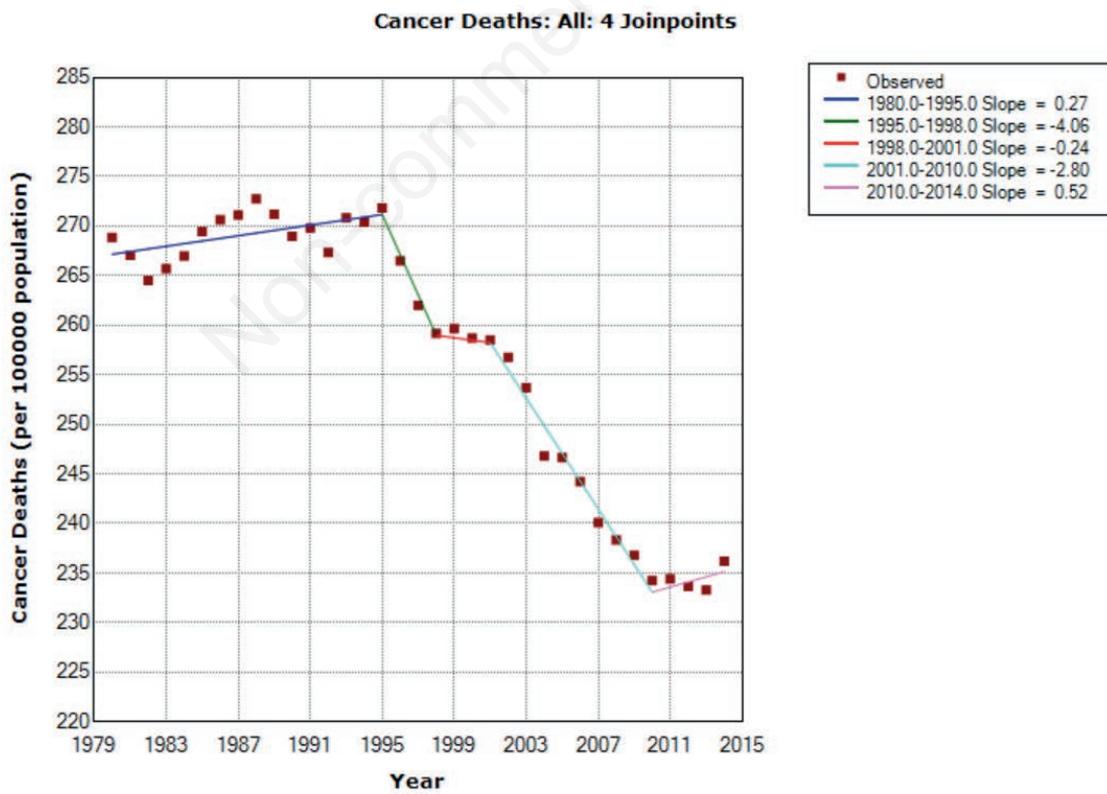


Figure 3. JOINPOINT analysis of cancer mortality (sum of all deaths due to the 29 cancers investigated) per 100,000 per population over the period 1980-2014.

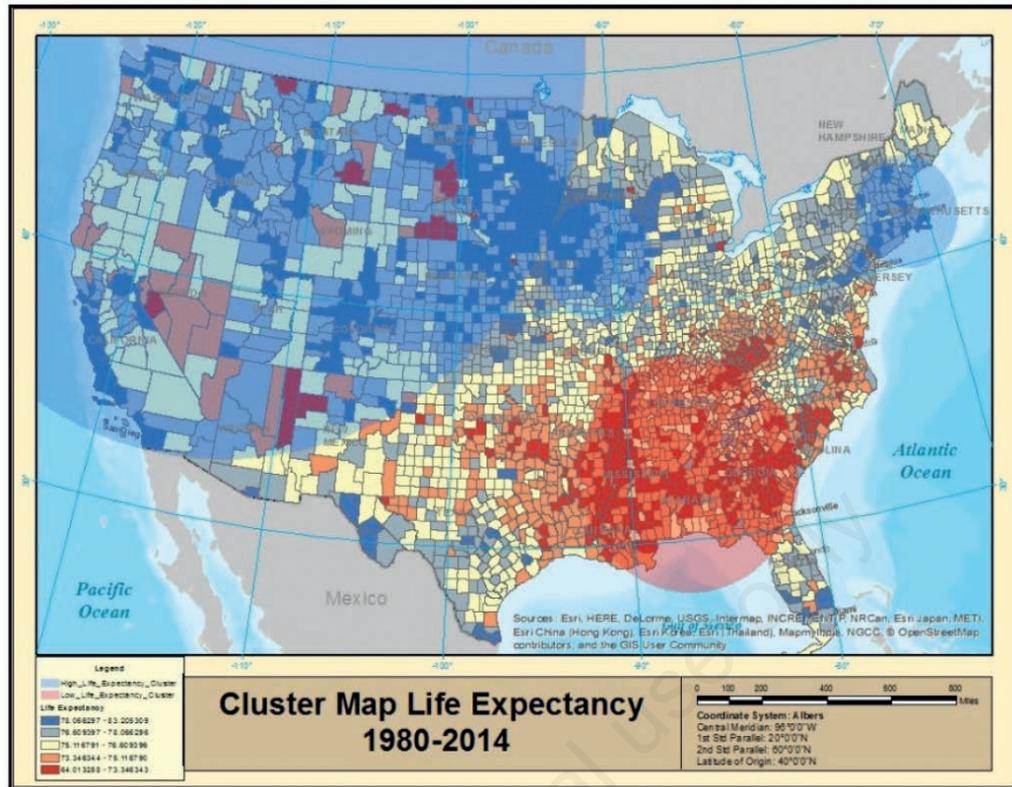


Figure 4. Map of purely spatial clusters of life expectancy over the period 1980-2014.

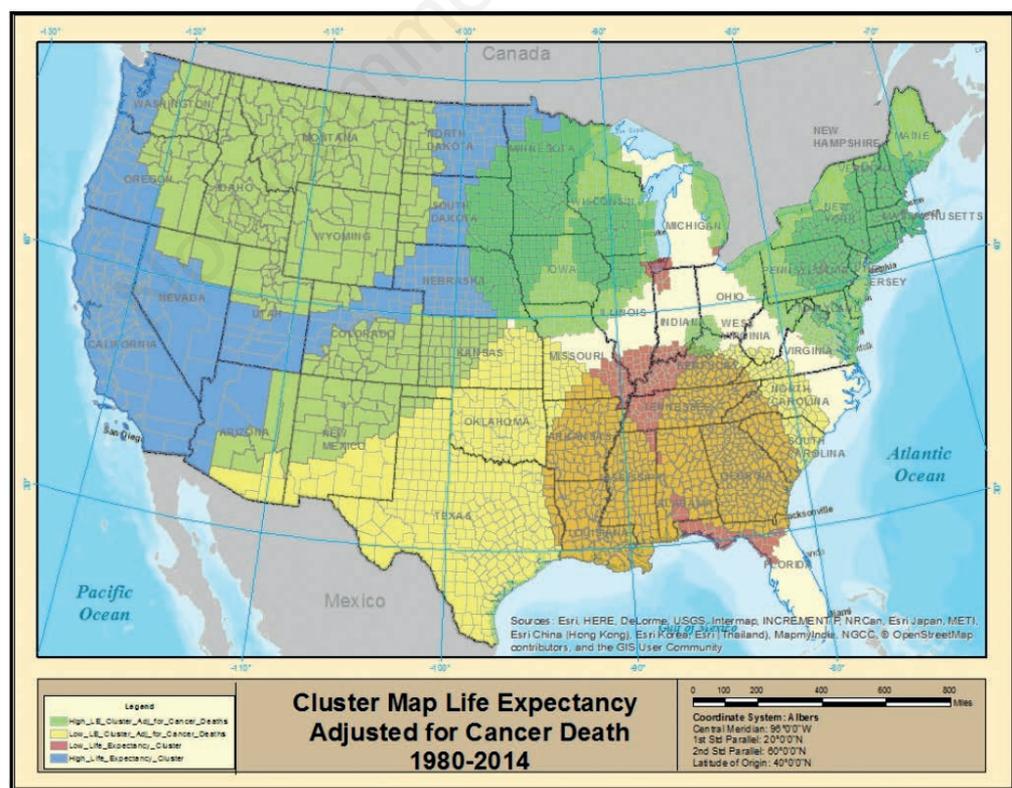


Figure 5. Map of purely spatial clusters of life expectancy adjusted for cancer mortality over the period 1980-2014.



Figure 6. Map of purely spatial clusters of life expectancy adjusted for cardiovascular disease mortality over the period 1980-2014.

over time from $r=-0.75$ in 1980 to $r=-0.87$ in 2014, with the correlation between life expectancy and CVD mortality changing from -0.78 in 1980 to $r=-0.87$ in 2014. While the US averages may show a certain level of worsening over the past 35 years, the spatial cluster maps make it clear where in the USA these two causes of death are significantly associated with life expectancy being very high or very low. Our research questions have been addressed and we can conclude that: i) there are significant, purely spatial clusters in some geographical areas where life expectancy rates are significantly higher (or lower) than in the rest of the contiguous US; ii) there is a significant association between the life expectancy level and the corresponding CVD mortality, and there is also a significant association between the life expectancy level and the corresponding overall cancer mortality; iii) the trends (regression slopes) over time for the USA in life expectancy mortality, CVD mortality, and cancer mortality have significantly changed after 2009-2010.

This surveillance study is providing useful information on the mortality due to CVD and on the overall cancer mortality on 29 cancer types for the years 1980-2014, in addition to identifying the associations between life expectancy, CVD mortality and cancer mortality. By 2009-2010, the life expectancy appears to be declining, and the mortality rates for all cancers and CVD increased.

References

- Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, Bergeron A, Cutler D, 2016. The Association between Income and Life Expectancy in the United States, 2001-2014. *JAMA* 315:1750-66.
- Deaton A, Lubotsky D, 2003. Mortality, inequality and race in American cities and states. *Social Sci Med* 56:1139-53.
- Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, Morozoff C, Kutz M, Huynh C, Barber R, Shackelford K, Mackenbach JP, Lenthe FJ, Flaxman AD, Naghavi M, Mokdad A, Murray CJ, 2016. US County-level trends in mortality rates for major causes of death, 1980-2014. *JAMA* 316:2385-401.
- Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, Morozoff C, Mackenbach JP, van Lenthe FJ, Mokdad AH, Murray CJL, 2017. Inequalities in life expectancy among US Counties, 1980 to 2014: temporal trends and key drivers. *JAMA Int Med* 2017;177:1003-11.
- Ezzati M, Horwitz ME, Thomas DS, Friedman AB, Roach R, Clark T, Murray C, Honigman B, 2012. Altitude, life expectancy and mortality from ischaemic heart disease, COPD and cancers: national population-based analysis of US counties. *J Epidemiol Commun Health* 66:e17.

- Kaplan GA, Pamuk ER, Lynch JW, Cohen RD, Balfour JL, 1996. Inequality in income and mortality in the United States: analysis of mortality and potential pathways. *BMJ* 312:999.
- Kim HJ, Fay MP, Feuer EJ, Midthune DN, 2000. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med* 19:335-51.
- Kulldorff MA, 1997. Spatial scan statistic. *Commun Stat Theory Methods* 26:1481-96.
- Kulldorff M, Information Management Services Inc, 2009. SaTScan™ v8.0: Software for the spatial and space-time scan statistics. Available from: <http://www.satscan.org/>
- Kulldorff M, Huang L, Konty K, 2009. A scan statistic for continuous data based on the normal probability model. *Int J Health Geogr* 8:58.
- Meara ER, Richards S, Cutler DM, 2008. The gap gets bigger: changes in mortality and life expectancy, by education, 1981-2000. *Health Affairs* 27:350-60.
- Mokdad A, Dwyer-Lindgren L, Fitzmaurice C, Stubbs RW, Bertozzi-Villa A, Morozoff C, Charara R, Allan C, Naghavi M, Murray CJL, 2017. Trends and patterns of disparities in cancer mortality among US counties, 1980-2014. *JAMA* 317:388-406.
- Olshansky SJ, Antonucci T, Berkman L, Binstock RH, Boersch-Supan A, Cacioppo JT, Carnes BA, Carstensen LL, Fried LP, Goldman DP, Jackson J, Kohli M, Rother J, Zheng Y, Rowe J, 2012. Differences in life expectancy due to race and educational differences are widening, and many may not catch up. *Health Affairs* 31:1803-13.
- Pope CA, Ezzati M, Dockery DW, 2009. Fine-particulate air pollution and life expectancy in the United States. *N Engl J Med* 360:376-86.
- Tarkiainen L, Martikainen P, Laaksonen M, Vlakonen T, 2012. Trends in life expectancy by income from 1988 to 2007: decomposition by age and cause of death. *J Epidemiol Commun Health* 66:573-8.
- Wang L, Wei B, Li Y, Li H, Zhang F, Rosenberg M, Yang L, Huang J, Krafft T, Wang W, 2014. A study of air pollutants influencing life expectancy and longevity from spatial perspective in China. *Sci Total Environ* 487:57-64.

Non-commercial use only