Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 555–573
http://www.ssca.org.in/journal

Analysis of Spatial and Temporal Patterns in Deaths of Despair in the Appalachian Region of the United States

Vishal Deo[1,2†], Raanan Gurewitsch[3†], Saurav Guha[2,4], Meghana Ray[2] and Saumyadipta Pyne[2,5]

[1] *National Institute for Research in Digital Health and Data Science, ICMR New Delhi, India*
[2] *Health Analytics Network, Pittsburgh, PA, USA*
[3] *University of Pittsburgh, Pittsburgh, PA, USA*
[4] *Department of Statistics, Mathematics & Computer Application, Bihar Agricultural University, Bhagalpur, Bihar, India*
[5] *Department of Statistics and Applied Probability, University of California Santa Barbara, Santa Barbara, CA, USA*

[†] These authors contributed equally.

Received: 02 August 2024; Revised: 28 September 2024; Accepted: 04 October 2024

Abstract

Mortality data in the United States (U.S.) revealed a precipitous rise in deaths among non-Hispanic white populations starting in the later part of 1990s due to such causes as Suicide, Alcohol consumption, and Drug accidental overdose. In particular, opioid-related deaths have increased dramatically across the U.S. during this period. For a systemic analysis of the temporal and spatial patterns of this critical phenomenon from the perspective of public health, we studied it in the context of Appalachian Region (AR), which spans across 13 U.S. states and is home to more than 8 percent of the country's population. We identified 8 spatial and temporal metaclusters of AR counties with relatively high rates of deaths due to the above-mentioned causes over the period 1979-2017 based on U.S. county- and cause-specific mortality data. Thus, we analyzed the mortality trends for each of the metaclusters, which were characterized based on their respective demographic and socioeconomic changes.

Key words: Deaths of despair; Mortality data; Spatial clustering, Temporal patterns; Opioid epidemic; Appalachian Region.

AMS Subject Classifications: 62H11, 62H30

Corresponding Author: S. Pyne
Email: spyne@ucsb.edu

For successful policy-making,
a government needs good statistics
as well as good statisticians.
One is not a substitute for the other.

Calyampudi Radhakrishna (C.R.) Rao

1. Introduction

Advancements in science, medicine and technology have extended life expectancy and improved health outcomes over the past half-century in the U.S., but not everyone has benefited equally. Researchers studying mortality have noted a precipitous rise in deaths between 1999 and 2005 due to such causes as Suicide, Alcohol consumption and Drug accidental overdose. Here, these causes of deaths are collectively denoted by the acronym: *SAD*. This phenomenon was characterized initially among non-Hispanic white Americans – mostly without a four-year college degree – as "deaths of despair" by Anne Case and Angus Deaton (Case and Deaton, 2015). They noted that white working-class lives over the last half century were affected by long-term labor market declines (Case and Deaton, 2020) which, in turn, led to a decline of families and relationships, limited access to high-quality healthcare, increased social isolation and loneliness, and a general loss of hope for the future (George *et al.*, 2021). Simultaneously, increases in ease of access to handguns, inexpensive alcohol, and prescription or non-prescription drugs including opioids have played their role in these excess, premature, and sometimes preventable, SAD deaths (George *et al.*, 2021; Shiels *et al.*, 2020).

Notably, while the historic deaths of despair were largely observed among non-Hispanic whites in rural America, more recently, around 2015, they increased across all races (Hedegaard *et al.*, 2018b,a). In 2017, there were 158,000 documented despair-related deaths that contributed to the longest sustained decline in life expectancy since 1915 (Woolf *et al.*, 2018). In the period between 2000 and 2017, there were 1,446,177 drug poisoning, suicide, and alcohol-induced premature deaths in the U.S., that included 563,765 drug poisoning deaths (17.6 per 100,000 person-years), 517,679 suicides (15.8 per 100,000 person-years), and 364,733 alcohol related deaths (10.5 per 100,000 person-years). These amounted to 451,596 excess deaths than those expected based on the rates of 2000 (Shiels *et al.*, 2020). For instance, alcohol related deaths increased by 77% from 2000 (19,627) to 2016 (34,857). Alcohol deaths have risen in all races/ethnicities and across all age groups in both men and women between 2000 to 2016 (Spillane *et al.*, 2020).

Opioid-related deaths have increased dramatically in the past two decades in the U.S. In 2016 alone, there were 45,838 opioid related deaths, and in 2017, the U.S. Department of Health and Human Services declared the opioid epidemic as a "public health emergency" (U.S. Department of Health and Human Services, 2023). Increase in mortality due to drug overdose rose by 15%, while alcohol-related and suicide deaths have increased yearly by 4.1% and 1.5% respectively (Shiels *et al.*, 2020). Trajectories in the former rates of mortality have also varied by geographical regions – high in some predominantly rural states such as Maine, Kentucky and West Virginia, with the lowest in other largely rural states such as

Nebraska and Iowa (Rigg *et al.*, 2018). The distribution of local patterns of the phenomenon is complex, and national-level analyses (*e.g.*, Jalal *et al.* (2018)) seldom consider myriad key factors that may influence the data. These include different types of opioids (say, cheaper synthetic ones, *e.g.*, Fentanyl), societal stigma (which in turn influences treatment seeking behaviors), differences in reporting or physician training, accuracy of determination of the cause of death, effectiveness of different interventions in addressing the specific types of opioid use, *etc.*

Discussions that focus on opioid mortality often overlook the intersectionality among its various social determinants including educational attainment or employment dynamics of a population. The effects of such factors are both spatial as well as temporal in nature. Taking cognizance of the fact that despair is, in general, a complex psychosocial phenomenon, we chose to focus our study specifically on SAD as the notified causes of death (as given by the ICD-10 codes) in the mortality database. Attempts were made in the past to broadly study national level data on socioeconomic disparities, mortality statistics, or the opioid epidemic (Wallace *et al.*, 2019; Case and Deaton, 2021; Jalal *et al.*, 2018). However, for gaining key insights into SAD deaths, we think it is more effective to concentrate on the occurrence (or recurrence) of mortality patterns in a particular highly-affected geographical region, which could then be partitioned into spatial and temporal subregions for systematic investigation. Towards this, in the present study, we focused on the Appalachian Region (AR) that is known for high rates of poverty and mortality due to despair as well as being among those parts of the U.S. that were seriously affected by drug overdose deaths over the past few decades (Rigg *et al.*, 2018; NACo and ARC, 2019).

Notably, AR is a 205,000-square-mile region that spans the Appalachian Mountains from southern New York to northern Mississippi. It includes all of West Virginia (WV) and parts of 12 other states: Alabama (AL), Georgia (GA), Kentucky (KY), Maryland (MD), Mississippi (MS), New York (NY), North Carolina (NC), Ohio (OH), Pennsylvania (PA), South Carolina (SC), Tennessee (TN), and Virginia (VA). AR includes 423 counties (around 13% of all counties in the U.S.) and 8 independent cities in 13 states, and has a population of approximately 25 million people. It is divided into 5 sub-regions: Northern, North Central, Central, South Central, and Southern. As per the Appalachian Regional Commission (ARC), the region has overall low educational attainment, increasing unemployment and poverty. In 2019, 24.9% and 24.6% of eligible individuals earned Bachelor's degrees in mining and non-mining counties respectively compared to 32.8% in the rest of the U.S. (Bowen *et al.*, 2020). Between 2005 and 2020, employment in the coal industry fell by around 54% (Bowen *et al.*, 2020). Between 2013 and 2017, poverty rates in Appalachia averaged 16.3% compared to 14.6% for rest of the U.S. A regional analysis indicates that poverty rates ranged between 6.5% to 41% with poverty mostly concentrated in central Appalachia encompassing Eastern KY and WV (Appalachian Regional Commission, 2019).

AR has seen a steeper rise in deaths of despair since around 1998, especially among the middle-aged population, as compared to the non-Appalachian regions of the country (Meit *et al.*, 2017). While the region is home to around 32.5% of the U.S. population, it accounted for 49.6% of excess deaths in U.S. caused by the increase in midlife mortality during 2010-2017 (Meit *et al.*, 2019; Woolf *et al.*, 2019). Further, Meit *et al.* (2019) have noted that this disparity in deaths of despair is more evident in the Central and North Central Appalachian sub-regions. Accidental drug overdoses were identified as a major contributor towards the rising deaths of despair in AR between 1999 and 2017, sustained by the easy and abundant availability of prescription opioids and heroin in the region (Woolf *et al.*, 2019; Monnat, 2020). In rural Appalachia, women have been found to be at a higher risk of committing suicide than men (Christine *et al.*, 2020). Declining manufacturing and mining industries, persistent poverty, rurality, social isolation, and physically demanding and injury-prone manual labor jobs have been studied as some of the possible socio-economic determinants of distress in AR, *e.g.*, George *et al.* (2021), Rigg *et al.* (2018), Meit *et al.* (2017), Meit *et al.* (2019), Woolf *et al.* (2019), and Monnat (2020).

In this study, our aim is to identify sub-regions of AR with high prevalence of SAD deaths at multiple time-periods, and to investigate the association of SAD mortality trends in these sub-regions with their economic and demographic characteristics. Given the dynamic nature of such characteristics at local (county) levels, we divided the overall study time-period of 1979-2017 into eight five-year periods, and identified flexibly-shaped spatial clusters of AR counties based on high SAD mortality rates for each period. Further, 8 metaclusters were constructed by combining spatially contiguous counties that had multiple occurrences among the clusters identified in different time-periods. These 8 metaclusters represent sub-regions with persistent prevalence of high SAD mortality in AR, which were then characterized based on relevant covariates. The metaclusters were compared with respect to temporal trends in various demographic and economic parameters such as annual average overall employment rate, industry-specific employment rates (mining and manufacturing), population size, median age, and median household income. After description of the methods and results in the subsequent sections, we end with an overall discussion of the analysis, its findings and limitations.

2. Data

We obtained the time-series data of age-adjusted mortality rates due to the three SAD causes (based on the corresponding ICD 10 codes) for each county in AR from the publicly accessible Mortality Information and Research Analytics System (MOIRA) of University of Pittsburgh (www.moira.pitt.edu). MOIRA data is sourced from the Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS), and the U.S. Census Bureau. The MOIRA system facilitates extraction and visualization of U.S. mortality and population data in a standardized format and categorized by causes of death given by International Classification of Diseases (ICD 10) codes. Data was collected for the period 1979-2017, which also contained mortality rates grouped by sex and race. Additional data such as on employment, wages, population size, median age, and median household income was obtained from the official websites of the U.S. Bureau of Labor Statistics (www.bls.gov) and the U.S. Census Bureau (www.data.census.gov).

3. Methods

For each of the eight five-year periods, we performed spatial clustering of the 423 Appalachian counties based on county-level SAD Age-Adjusted mortality Rates (AAR) using a flexibly shaped spatial scan statistic due to Tango and Takahashi (2005) implemented with a restricted likelihood ratio in the R package rflexscan (Otani and Takahashi, 2021). The counties appearing in these spatial clusters were identified and ranked by their number of occurrences over the eight time-periods. Recurrent counties, *i.e.*, counties with more than one occurrence in the time-period-specific clusters, were combined using the K (=1) nearest-neighbor strategy to obtain the final spatiotemporal metaclusters. We characterized the metaclusters using known Socio-Economic Status (SES) and race-based county labels (Wallace *et al.*, 2019). Smooth log-transformed trends of SAD AAR were obtained using the MortlitySmooth package in R (Camarda, 2012) and were plotted by age, sex, race, and SAD causes of death for each of the 8 metaclusters. SAD AARs were also predicted using the same library beyond 2017 for each metacluster until 2020, to avoid conflation with the Covid-19-associated mortality rates of the same areas thereafter.

Trends of annual employment rate in the mining and manufacturing industries were also plotted for the metaclusters. In addition, distributions of average annual unemployment rates in the metaclusters in five-year periods were visualized using boxplots. Percentage changes in population size and median age of the population of the metaclusters from 1980 to 2020 were also evaluated for the metaclusters. Total population of metaclusters in a year was calculated as the sum of census population of the counties. Median age of the metaclusters for a given year was calculated as the median of the county-wise median age. In addition, median household income of the metaclusters was calculated as the median of the county-wise median household income. Since the county-wise estimates of median household income were available for the years 1979 and 2021 (which are based on the 1980 and 2020 census, respectively), the percentage change was calculated from 1979 to 2021. All the three metacluster-wise percentage changes were obtained as the median of the percentage changes calculated for the respective counties during the mentioned period.

4. Results

Spatially-flexible scan statistics identified clusters comprising of the counties in AR with relatively higher SAD age-adjusted mortality rates (AAR) for each of the 8 five-year time-periods (Figure 2). Although the compositions of the clusters vary across time-periods, some counties appeared recurrently in the identified clusters over time. Such counties are mostly concentrated around the South Central, Central, and North Central Appalachia, with a few in Northern and Southern regions. Assuming that a higher number of recurrences of any county in the clusters over the eight time periods would indicate a longer prevalence of SAD mortality therein, we used the K (=1) nearest-neighbor strategy for selecting nearby counties that have multiple such recurrences, and then combining them to form a *metacluster*. The neighborhood of (two or more) counties is determined by their sharing of common boundary lines or points even if they lie across different states.

The above procedure led to the construction of 8 metaclusters of counties spanning across AR that are both spatially contiguous and temporally recurrent hotspots of SAD deaths (Figure 3). Five of these eight metaclusters are located around the western regions of the Central and South-Central Appalachia, with one cluster extending to the lower region of the North Central Appalachia. Among the remaining three clusters, which are relatively smaller in size, two are located in the Northern Appalachia and one in the Southern Appalachia. Details of the identified metaclusters are provided in Table 1. A closer look at the physical map of the Appalachian region (Figure 3 inset) reveals the location and the underlying landform patterns of these metaclusters. Evidently, the metaclusters are located mostly along the Valley & Ridge region, and the southern part of the Blue Ridge Mountains.

Figure 1: Details of spatiotemporal metaclusters

Cluster No.	Cluster Name	State(s) of counties	No. of counties	County-FIPS	Appalachian Subregion (no. of counties)
1	Alabama (AL)	Alabama	18	37005, 37067, 37171, 37197, 51021, 51027, 51051, 51063, 51077, 51089, 51105, 51141, 51155, 51167, 51169, 51185, 51195, 51197	Central (7), and South Central (11)
2	Eastern PA	Pennsylvania	15	13241, 37021, 37027, 37043, 37087, 37089, 37099, 37113, 37149, 37161, 37173, 37175, 45021, 45073, 45083	South Central (11), and Southern (4)
3	Kentucky (KY)	Kentucky	12	21013, 21025, 21071, 21095, 21115, 21119, 21131, 21133, 21189, 21193, 21195, 21235	Central (12)
4	South NC + SC + GA	North Carolina, South Carolina, and Georgia	12	47013, 47001, 47025, 47049, 47057, 47059, 47063, 47093, 47145, 47151, 47163, 47173	Central (7), and South Central (5)
5	Tennessee (TN)	Tennessee	10	54005, 54019, 54039, 54045, 54047, 54055, 54059, 54081, 54089, 54109	North Central (5), and Central (5)
6	VA + North NC	Virginia and North Carolina	6	42003, 42007, 42051, 42125, 54009, 54029	Northern (6)
7	Western PA	Pennsylvania	5	42069, 42079, 42089, 42107, 42113	Northern (5)
8	West Virginia (WV)	West Virginia	4	1009, 1055, 1073, 1095	Southern (4)

To illustrate if spatiotemporal patterns of socioeconomic status (SES) have any association with the identified metaclusters of high SAD mortality rates, we used the SES class labels due to Wallace *et al.* (2019). The labels 1 & 8 represent high SES, 2 mid/low SES, and 4 mid SES; where 1 & 2 are semi-urban, 8 is rural, and 4 mostly-rural counties. Thus, Figure 4 provides us with a nuanced characterization of each metacluster. We observe that 6 out of 8 metaclusters have a majority of their counties falling in low SES categories. However, the two metaclusters in PA – Eastern PA and Western PA – appear to be distinctive from the rest as they have a more balanced distribution of both semi-urban and rural counties with high as well as mid SES. We discuss about this point further below. Here, we note that the categorization by Wallace *et al.* (2019), which is based on relatively recent SES of the counties, may not capture the full dynamics of SES over the entire time-period of this study.

Historical trends of SAD AAR (in logarithmic scale) in each of the 8 metaclusters from 1979 to 2017 for different causes of deaths of despair, age groups, races and sexes are presented in the Figures 5, 6, 7 and 8, respectively. The changing dynamics of the cause-specific SAD AARs in the 8 metaclusters is visualized in Figure 5. While the contribution of drug overdose deaths to SAD AAR was almost negligible in the early 1990s, it has grown

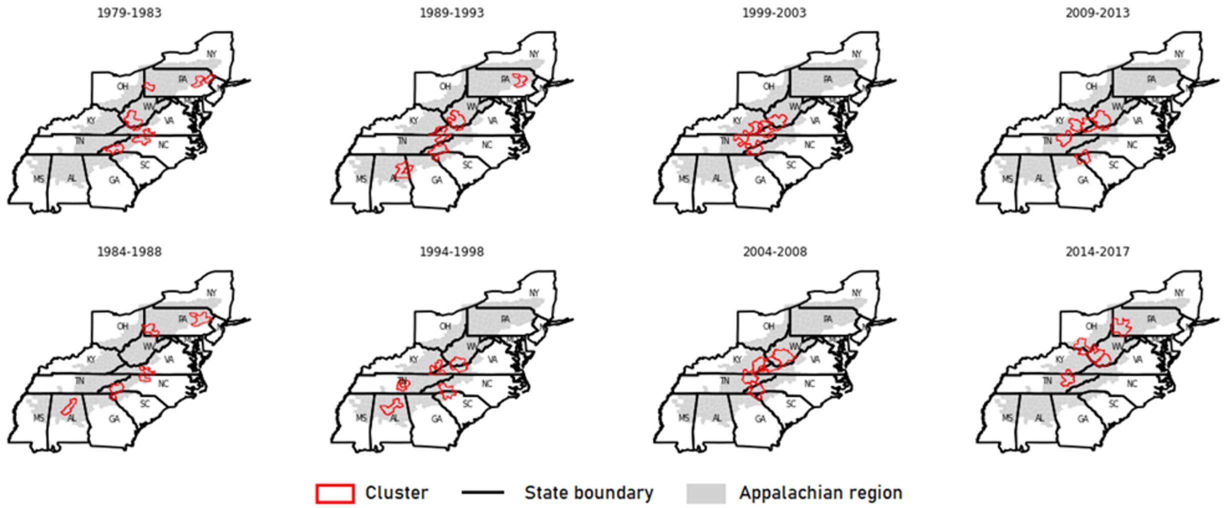

Figure 2: Spatial clustering of the AR counties in terms of their SAD AAR in each of the 8 successive 5-year time-periods between 1979 and 2017. The identified clusters' boundaries are shown with red lines. AR (grey area) and the state boundaries (black lines) are included for visual reference.

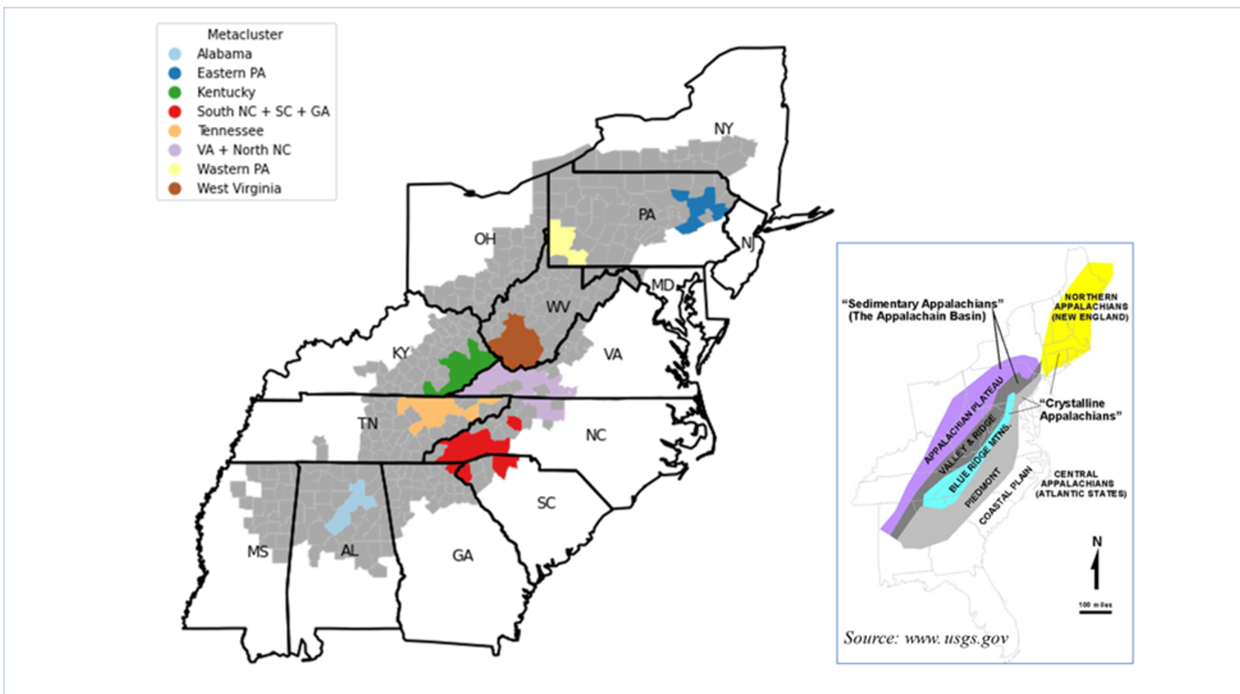

Figure 3: The 8 spatial metaclusters of SAD deaths in AR produced from the spatial clusters over time-periods between 1979 and 2017. These are shown in distinct colors and their labels in the legend. For visual reference, AR (grey area) and the state boundaries (black lines) are included along with an inset physical map of AR (due to www.usgs.gov).

continuously at a fast pace since then, and has become comparable to those due to suicides and alcohol related deaths. In fact, in some metaclusters (TN and VA + North NC), drug

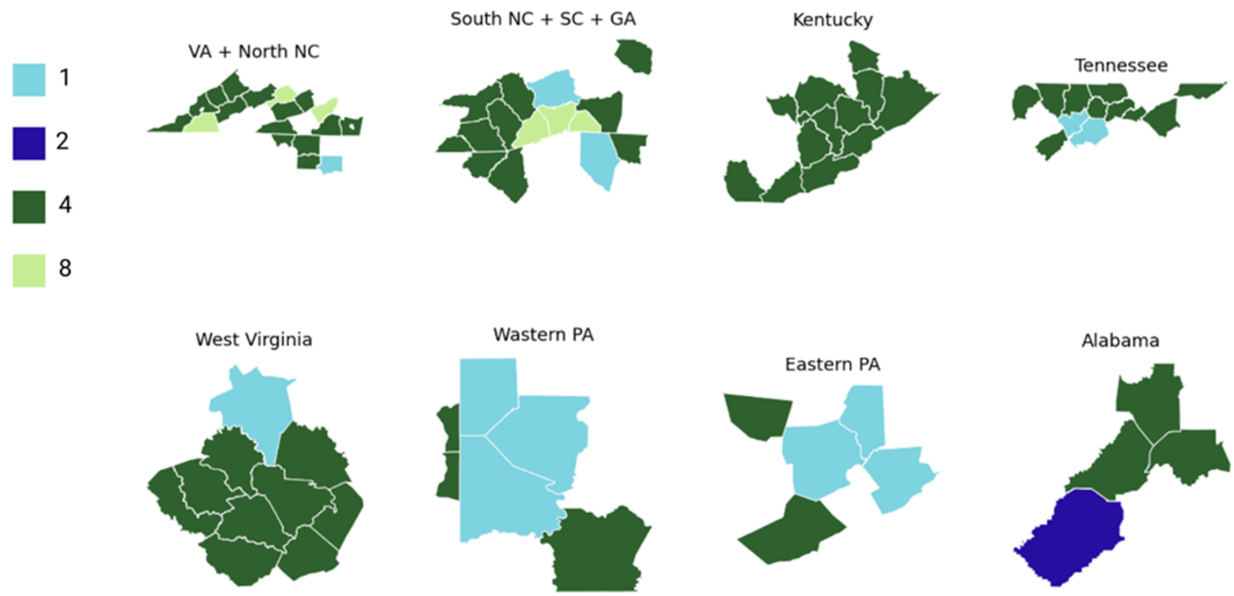

Figure 4: Characterization of the spatial metaclusters based on the known SES classification (given in the color-key) of their constituent counties.

overdose mortality AAR has even surpassed that of alcohol-related deaths and suicides. A slightly improved scenario on drug-overdose deaths could be noted in metacluster 3 (KY) by comparing the trends in Figures 5 and 6 during towards the end of the study period. Notably, the AAR trends of suicides and alcohol related deaths have remained high, and in fact, more or less constant, over the study period in all metaclusters.

Notably, the most alarming pattern appears in Figure 6 in which the SAD mortality AARs for both the younger and older age groups have increased over the years in all metaclusters; but the rate of change is markedly higher for the former age group of < 45 years. SAD AAR for the older age group (≥ 45 years) has been high since 1979, with occasional plateauing around the last decade of the 20th century, before it started increasing again, albeit at a slower pace compared to the younger age group. The trend for the younger age group is present in every metacluster, rising from around 12-20 per 100,000 in the year 1979 to around 33-55 per 100,000 in the year 2017, *i.e.*, at the end of the study-period. Interestingly, while the SAD AAR trends have continued to rise over the decades, Figure 7 shows little racial difference therein between whites and non-whites, except for marginally higher rates for the whites in the last two decades. Overall, the gradual and continual rise in the SAD AARs among the female and the younger populations – vis-a-vis the traditional trends of the male and the older populations – are clearly visible from Figures 6 and 8 respectively.

Moreover, the younger age group's SAD AAR has been continuously increasing in all metaclusters, except for metacluster 3 (KY) where a downward trend was observed in the last decade. In all other metaclusters, the gap between the SAD AAR of the younger and the older populations has continued to get narrower. Metacluster 6 (VA + North NC) has seen the sharpest rise in SAD mortality rate of the younger age group in recent years, with its value reaching almost 90 per 100,000. This phenomenon of alarmingly increasing SAD AAR among the younger age group is prevalent among all races, with almost similar trends for the non-Hispanic white and the other race groups (Figure 7). Although the SAD AAR

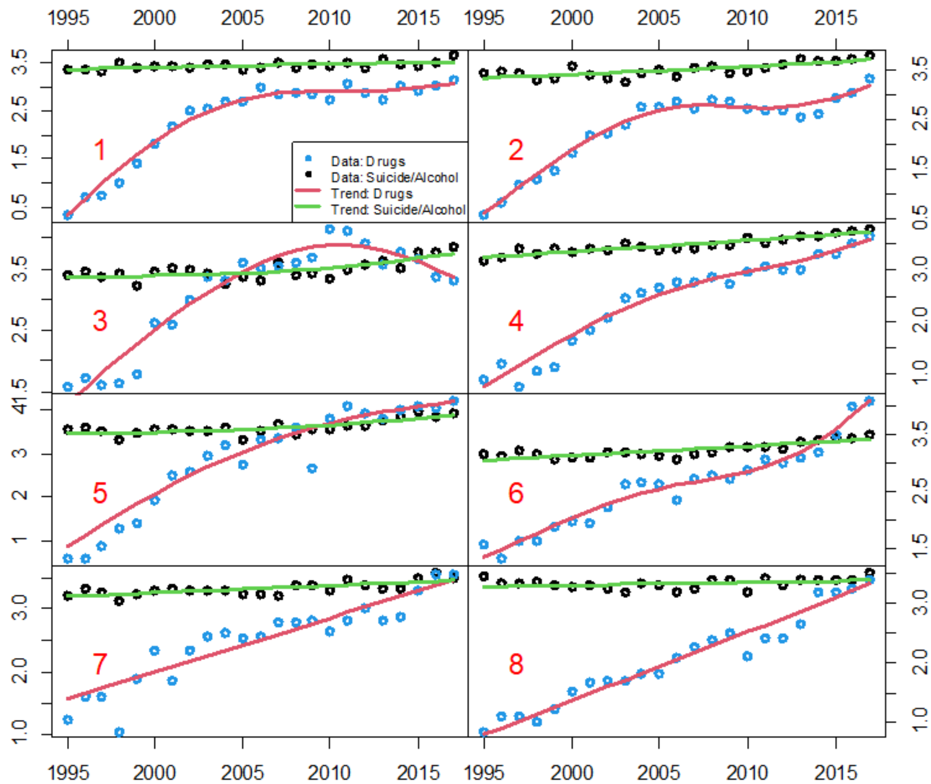

Figure 5: Longitudinal trends of SAD AAR for different causes of deaths of despair in the 8 spatial metaclusters of AR. The historically more common causes, *i.e.*, suicides and alcoholism (SA: green curves) are compared with the drug overdoses (D: red curves).

for younger age groups has been consistently high for the male population as compared to the female population, the rate of increase of the SAD AAR has been evidently higher for the female population in all metaclusters (see Figure 8).

For a visual comparison of the different metaclusters, Figure 9 overlays their trends of SAD AAR from 1979 to 2020 along with that of the remaining counties in AR (shown as a bold grey curve). Naturally, the latter, denoted by "Rest", has lower SAD AAR than every metacluster for every observed time-period while following a similarly increasing trend over time. In addition to this baseline trend, we also included the projected trend of SAD AAR for the time-period beyond 2017, up to 2020 (*i.e.*, the pre-pandemic years), as shown to the right of the dotted line. Overall, it is evident that the SAD AAR has been increasing in all metaclusters ever since 1979, and with a higher pace since around 2000. The projections provide clear insights into the temporal patterns for the metaclusters. For instance, the distinctive decline in SAD mortality of metacluster 3 (KY) in the last decade stands out among all of these trends. Metaclusters 1 (AL) and 5 (TN) started with very similar trends but went on to stray – during the 1990s – the most apart from each other. In fact, the latter is projected to have the highest SAD AAR among all the metaclusters exactly when the former is supposed to have AAR even lower than the Rest. Most importantly, several metaclusters {1, 2, 7, 8}, spanning various sub-regions of AR, that had exhibited different trends in the first decade, seemed to converge towards the Rest by the end of the study-period.

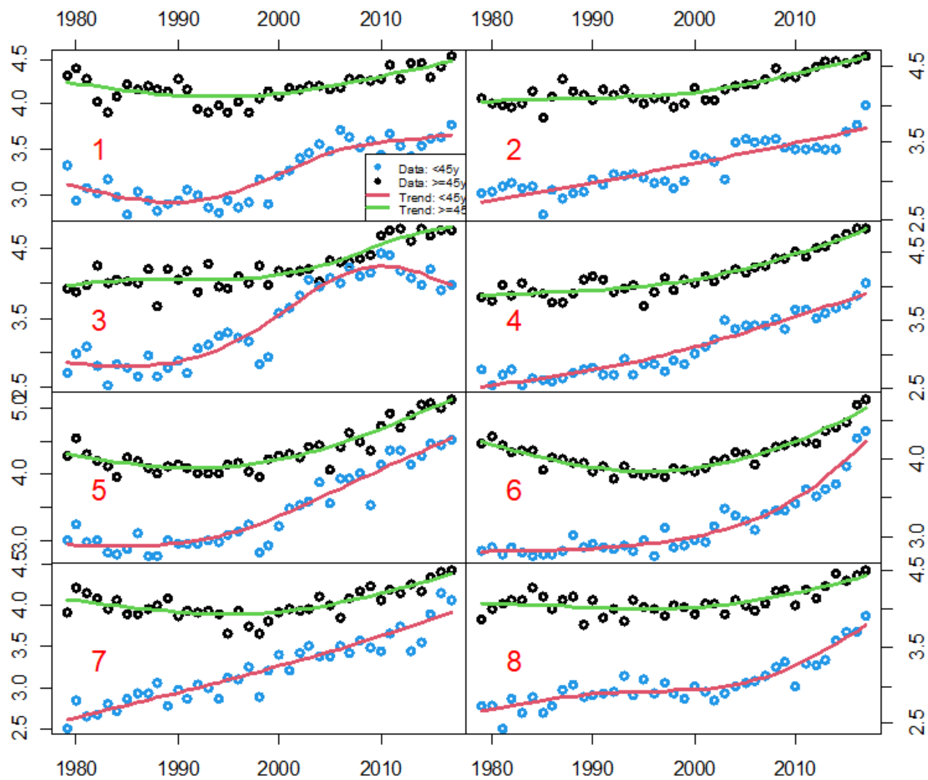

Figure 6: Longitudinal trends of SAD AAR for different age-groups in the 8 spatial metaclusters of AR. The older groups (≥ 45 years: green curves) are compared with the younger groups (<45 years: red curves).

To gain insights into the socioeconomic conditions of each metacluster, the unemployment rates of the counties in the eight metaclusters for each 5-year period from 1990 to 2020, are presented as boxplots in Figure 10. Although their scale and range of variation differ across the metaclusters, the trends are similar *e.g.*, the lowering of unemployment rates between 2000 and 2005, and then again around 2015. Comparing these trends with those of the SAD AAR (Figure 9), we can clearly observe sharp increases in the gradient of the SAD AAR around the years of higher unemployment rate. For instance, in the year 1995, unemployment rates of metaclusters 3 (KY), 4 (South NC + SC + GA), 5 (TN), and 7 (Western PA) were very high, and around the same time, sharp upward shift in the trend of SAD AAR can be observed for these metaclusters in Figure 10. However, there are no apparent reduction in the gradient of the trends of SAD mortality rate during periods of lower unemployment rates, possibly hinting at deeper structural reasons for despair that may not be fully mitigated by employment alone.

The heightened problem of unemployment in the AR over the past few decades is generally associated with the steady decline in mining and manufacturing industries in the region. Historical trends of annual average number of jobs in these industries in the 8 metaclusters are plotted against the national average in Figure 11. Clearly, the decline in such jobs has continued over the past five decades for every metacluster, despite a brief recovery in mining during 2010-2015. This could be attributed to a wide array of factors ranging from demographic changes in terms of aging and migration to economic drivers such

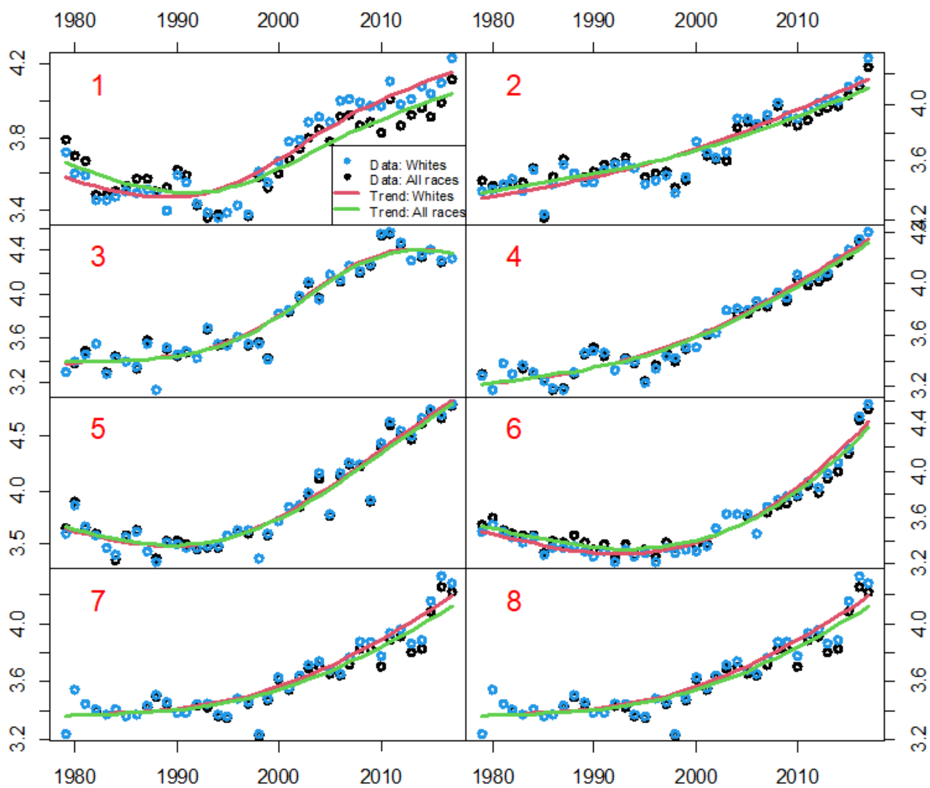

Figure 7: Longitudinal trends of SAD AAR for different races in the 8 spatial metaclusters of AR. The whites (red curves) are compared with the other races (green curves).

as outsourcing and global trade. Towards this, metacluster-wise median percentage change in population size, median age, and median household income, between 1980 and 2020 (Table 2) provide key insights into the shifts in the demographic and economic characteristics of each metacluster over the study-period. Interestingly, in 5 metaclusters, the population sizes have decreased between 5% and 29%. The same metaclusters, except for Western PA, are also characterized by higher rise in the median age of their population and much lower rise in their median household income as compared to the corresponding national change figures. The TN metacluster which has reported the highest SAD AAR in the last decade, has also seen the highest decline in population with the highest rise in median age and a decline in median household income.

Now we compare two metaclusters that are both from the same state, PA. The population of the Western PA metacluster decreased by 5% and its median age rose although lower than that of the U.S. Yet, its median household income has risen, in fact, by a greater percentage than the national average. This is likely to be driven by the urban sectors of the economy owing to Pittsburgh, the most prominent city in AR, in contrast to the ageing populations with limited opportunities for income generation among most of the other metaclusters that also have shrinking populations. Interestingly, it is also distinct from Eastern PA, which is one of the three metaclusters that have seen their populations rise. This metacluster has seen a 61% jump in population size and 51% increase in median household income, both of which are much higher than the national increments. These results not only

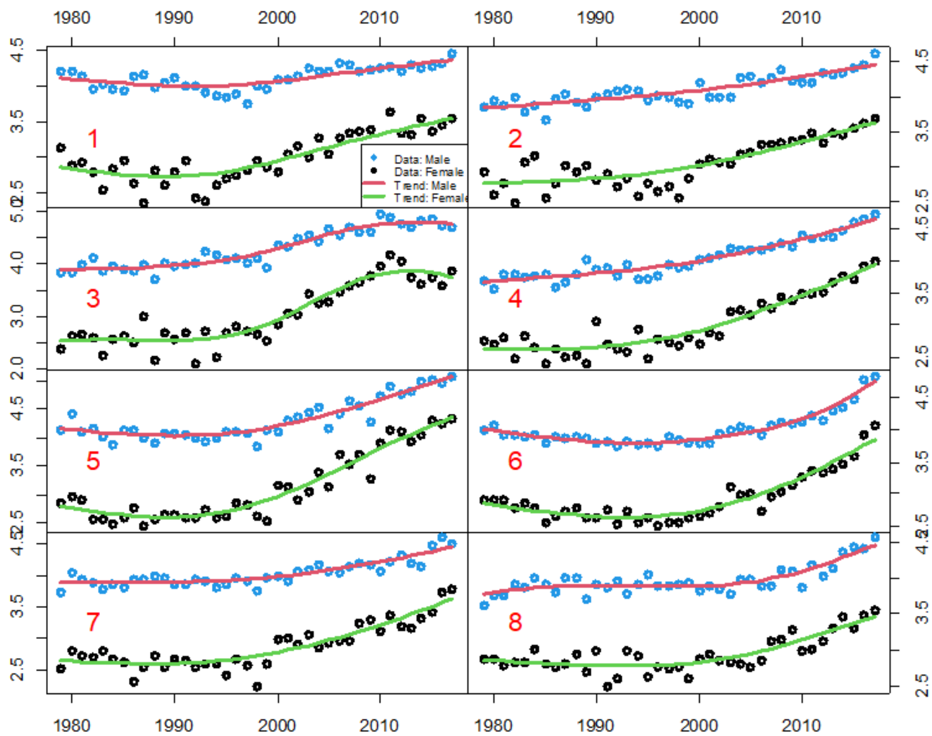

Figure 8: Longitudinal trends of SAD AAR for females (green curves) and males (red curves) in the 8 spatial metaclusters of AR.

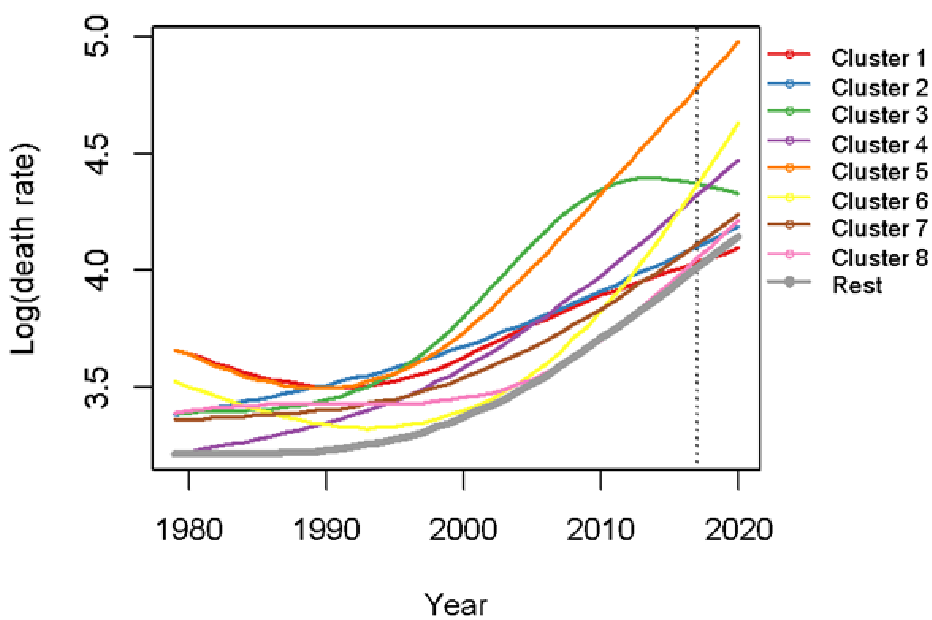

Figure 9: Trends of SAD AAR from 1979 to 2020 between 8 spatial metaclusters (as curves of different colors) are compared with the "Rest" of the Appalachian counties (bold grey curve). The predicted death rates for the time-period beyond 2017 are shown to the right of the dotted line.

showcase the distinct characteristics of these metaclusters even if they are from the same

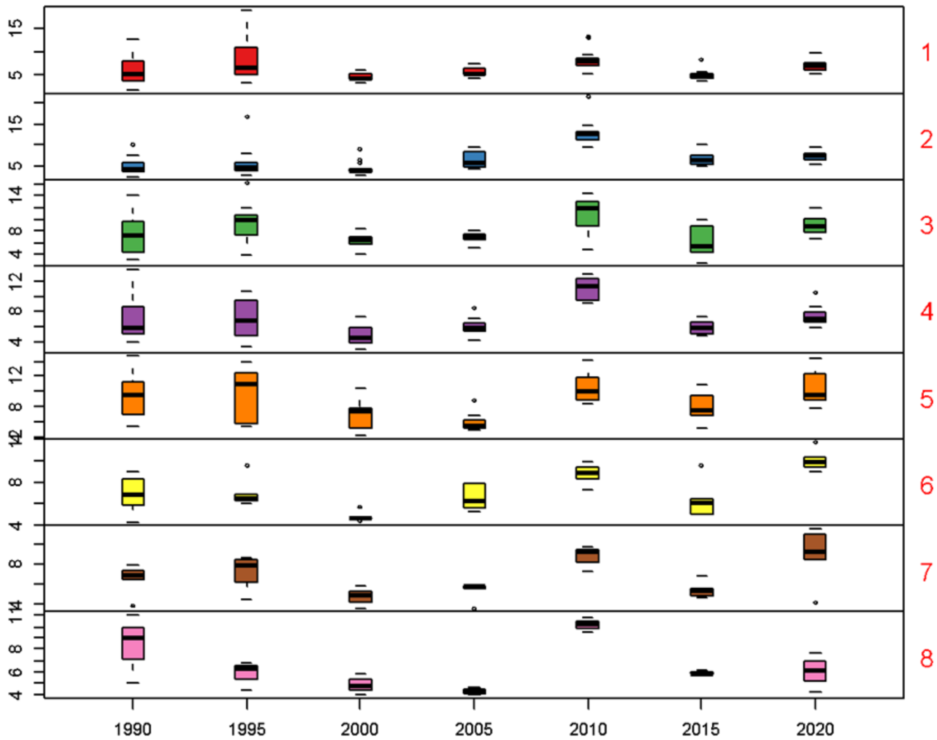

Figure 10: Unemployment rates (y-axis) over time (x-axis) are compared for the 8 spatial metaclusters (as labeled on the right) with boxplots for counties therein.

state, but also underscores how the complex problem of SAD deaths may not be adequately addressed merely by reducing poverty.

More granular insights can be derived from the county-level scatterplot of such changes in the demographic and economic parameters, upon grouping by the metaclusters, as shown in Figure 12. The inter-metacluster variation is more prominent in terms of the changes in population size. In general, counties with decline in population and steeper rise in median age have seen minimal rise in median household expenditure (smaller dots). Those with positive increases in population (appearing to the right-hand side of the dotted line) have witnessed rise in median household income (larger dots). As expected, higher rise in population is associated with lower rise in median age, but with some metaclusters that serve as notable exceptions. Moreover, we can observe intra-metacluster heterogeneity in terms of the changes in the demographic parameters and their interplay with household income. For example, some metaclusters have counties with moderate to high increases in their median age but notable rises in median household incomes. Such heterogeneities, within and across the metaclusters, may indicate the complexity underlying the phenomenon of SAD mortality, and underscore the need for investigating its social determinants at local community levels.

5. Discussion

Certain classic texts such as *The Other America* by Michael Harrington and *Night Comes to the Cumberlands* by Harry Caudill introduced Appalachian poverty to Americans during the early 1960s. The intense deprivation and hardships in AR led the then President

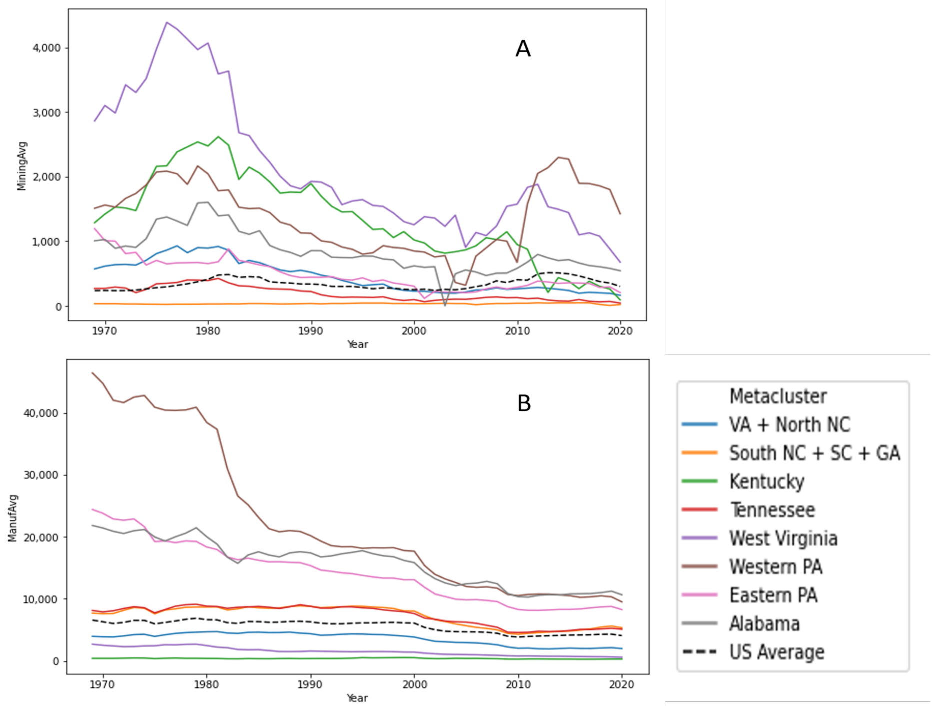

Figure 11: Longitudinal data on the average number of (A) mining and (B) manufacturing jobs in the 8 spatial metaclusters are shown (in different colors) against the national average (dotted).

John F. Kennedy to establish the President's Appalachian Regional Commission (PARC) in 1963. In its report, PARC categorically noted that "Appalachia is a region apart − both geographically and statistically" (PARC, 1964). In particular, AR "lags behind the rest of the Nation in its economic growth and that its people have not shared properly in the Nation's prosperity." In the subsequent decades, many steps have been taken towards reduction of the abject poverty in AR using various mechanisms, *e.g.*, the Appalachian Regional Development Act of 1965 (ARDA), which designated AR as a special economic zone and provided spending of more than \$23 billion. Six decades later, the observable and compelling phenomenon of SAD deaths makes it vital for the researchers to analyze patterns of such dire yet disparate outcomes that have persisted in certain areas – and even evolved during the opioid epidemic – against the complex socioeconomic background of AR.

In rural communities, residents are more likely to work in physically demanding and injury-prone job sectors such as farms, factories, and mines, as compared to their urban counterparts. These place workers at increased risks for chronic pain and disability (Keyes *et al.*, 2014). Between 2015 and 2019, the share of Appalachian residents who reported a disability was 16.2% compared to 12.6% for the U.S. (Pollard and Jacobsen, 2021). Indeed, the prevalence of midlife pain epidemic in the U.S., which was highlighted by Case and Deaton (2015), exacerbated by the surge in the use of prescription (or otherwise) painkillers since the mid-1990s, has well-documented links to both addiction and SAD deaths in AR (Quinones, 2015). Not surprisingly, therefore, many Appalachian communities that are mining-dependent became targets for heavy marketing of Oxycodone and other strong prescription opioids much earlier than the rest of the country (Rigg *et al.*, 2018). Detailed patterns of such substance

Figure 12: Percentage change in median age of the counties (grouped by their metacluster-specific colors) plotted against percentage change in their total population from 1980 to 2020. Size of the points is proportional to the percentage increase in the median household income of the counties.

and polysubstance uses among different population groups in the U.S. over the past 5 decades were identified by our previous studies based on NSDUH population surveys on substance use (Ray *et al.*, 2022).

In the present study, we identified 8 metaclusters of high rates of SAD deaths in AR over the period 1979-2017 based on U.S. county- and cause-specific mortality data. We observed patterns for each metacluster such as the dynamics for SAD mortality due to drug-overdoses, and its rising trends among the younger age group and women. We also noted heterogeneity among the metaclusters not only in terms of SES and rural/urban compositions but also their demographic and socioeconomic dynamics over the study-period. The patterns from our analysis showcase the need for further dissection of the data and covariates to detect the contributions of local vulnerabilities within each metacluster. For instance, the roles of such social determinants as poor public transport infrastructure that limits access to better jobs and healthcare, or sub-standard schools that may not prepare students for newer careers, cannot be ignored (George *et al.*, 2021). Based on our past use of subcounty-scale characteristics such as CDC Social Vulnerability Index, and techniques such as Small Area Estimation, we think further disaggregation of the SAD mortality data can lead to a more nuanced understanding of despair in the diverse communities of AR, and thus aid public health and policy-making (Stacy *et al.*, 2023). Structural solutions at local levels can address issues involving strategies that may go beyond even poverty and unemployment.

We understand that our study has certain limitations. While different approaches of space-time clustering are known (*e.g.*, Knox test), the one that we adopted here is based on our intention to avoid the identified clusters from being necessarily temporally contiguous. Therefore, to allow for the occasional "ups and downs" in the SAD mortality rates within a cluster, we first clustered the AR counties in each successive 5-year window, and then used their recurrence for constructing the metaclusters. Further, to allow flexibly shaped clusters, we decided not to use scan statistics based on a circular window (say, due to Kulldorff), which have difficulty in correctly detecting irregularly shaped clusters that are more realistic. Instead, we applied the flexible spatial scan statistic of Tango and Takahashi (2005), which is able to detect a cluster of any shape reasonably well as its relative risk increases during the Monte Carlo simulation used in this approach.

Figure 13: Metacluster-wise median % change in population size, median age, and median household income (in USD). Data Source: U.S. Census Bureau.

Metacluster No.	Metacluster Name	Total population			Median Age of Population			Median Household Income		
		1980	2020	Median % Change	1980	2020	Median % Change	1979*	2021	Median % Change
1	Alabama (AL)	746676	850344	-6%	31.4	46.9	48%	38229	45519	29%
2	Eastern PA	789917	1230264	61%	32.5	47.1	39%	37790	51817	51%
3	Kentucky (KY)	383919	297485	-27%	27.8	42.9	55%	30189	34474	14%
4	South NC + SC + GA	805216	1057295	27%	30.8	44.2	44%	32730	48986	53%
5	Tennessee (TN)	670198	485884	-29%	29.1	44.6	57%	39926	40591	-1%
6	VA + North NC	2103297	1808600	-19%	32.8	46.0	40%	52830	58447	17%
7	Western PA	807375	858706	-5%	35.2	43.4	24%	39372	58645	54%
8	West Virginia (WV)	876509	934903	25%	31.5	40.2	29%	36554	53034	35%
US		226,542,250	331,449,281	46%	30	38.8	29%	47396	69,717	47%

*Converted to 2021 USD

Dedication

We dedicate this paper to the memory of the legendary statistician, the late Professor C.R. Rao (1920-2023). Following his retirement from the Indian Statistical Institute, Dr. Rao had had a second illustrious academic career in the U.S.; in particular, at the University of Pittsburgh and the Pennsylvania State University. Incidentally, both of these institutions are located in AR, the region of focus in the present study. As the quote at the beginning of the paper underscores, Dr. Rao had a profound interest in the use of statistics for policy-making and public health (Rao *et al.*, 2017a,b). While his work was extended by us to address some recent public health problems (*e.g.*, Guha *et al.* (2022)), his longevity provided us with a direct historical connection to the classical past of statistics and its luminaries such as R.A. Fisher and P.C. Mahalanobis. One of the authors (SP) had the privilege of having Dr. Rao as a colleague at his eponymous institute in Hyderabad, India.

As a tribute to this trailblazing statistical scientist and an outstanding and prolific author as well as a wise and witty mentor to many, we echo the sentiment expressed in the Proceedings of the (U.S.) National Academy of Sciences earlier this year (DasGupta, 2024), “Goodbye, Dr. Rao. Thank you for your inspiration and guidance. We will remember you.”

Acknowledgements

Initial parts of the work was conducted with support from the Public Health Dynamics Laboratory at University of Pittsburgh. The authors declare no conflicts of interest.

References

Appalachian Regional Commission (2019). Poverty rates in Appalachia 2013-2017. https://www.arc.gov/map/poverty-rates-in-appalachia-2013-2017/.

Bowen, E., Christiadi, D. J., and Lego, B. (2020). An overview of coal and the economy in Appalachia. Appalachian Regional Commission.

Camarda, C. G. (2012). Mortalitysmooth: An R package for smoothing poisson counts with P-splines. *Journal of Statistical Software*, **50**, 1–24.

Case, A. and Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 15078–15083.

Case, A. and Deaton, A. (2020). *Deaths of Despair and the Future of Capitalism*. Princeton University Press.

Case, A. and Deaton, A. (2021). Life expectancy in adulthood is falling for those without a BA degree, but as educational gaps have widened, racial gaps have narrowed. *Proceedings of the National Academy of Sciences of the United States of America*, **118**, e2024777118.

Christine, A. S., Lancaster, K. E., Gaynes, B. N., Wang, V., Pence, B. W., Miller, W. C., and Go, V. F. (2020). The opioid and related drug epidemics in rural Appalachia: A systematic review of populations affected, risk factors, and infectious diseases. *Substance Abuse*, **41**, 69–69.

DasGupta, A. (2024). C. R. Rao: Paramount statistical scientist (1920 to 2023). *Proceedings of the National Academy of Sciences*, **121**, e2321318121.

George, D., Snyder, B., Van Scoy, L., Brignone, E., and et al (2021). Perceptions of diseases of despair by members of rural and urban high-prevalence communities. *JAMA Network Open*, **4**, e2118134.

Guha, S., Petersen, A., Ray, S., and Pyne, S. (2022). On Rao’s weighted distributions for modeling the dynamics of wildfires and air pollution. In Bapat, R. B., Karantha, M. P., Kirkland, S. J., Neogy, S. K., Pati, S., and Puntanen, S., editors, *Applied Linear Algebra, Probability and Statistics*, Indian Statistical Institute Series, pages 379–394. Springer.

Hedegaard, H., Curtin, S. C., and M, W. (2018a). Suicide mortality in the United States, 1999-2017. *NCHS Data Brief*, **330**, 1–8.

Hedegaard, H., Miniño, A. M., and Warner, M. (2018b). Drug overdose deaths in the United States, 1999-2017. *NCHS Data Brief*, **329**, 1–8.

Jalal, H., Buchanich, J. M., Roberts, M. S., Balmert, L. C., Zhang, K., and Burke, D. S. (2018). Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science*, **361**, aau1184.

Keyes, K. M., Cerda, M., Brady, J. E., Havens, J. R., and S, G. (2014). Understanding the rural-urban differences in nonmedical prescription opioid use and abuse in the United States. *American Journal of Public Health*, **104**, E52–E59.

Meit, M., Heffernan, M., and Tanenbaum, E. (2019). Investigating the impact of the diseases of despair in Appalachia. *Journal of Appalachian Health*, **1**, 1–18.

Meit, M., Heffernan, M., Tanenbaum, E., and Hoffman, T. (2017). Appalachian diseases of despair. The Walsh Center for Rural Health Analysis at NORC at the University of Chicago. https://www.arc.gov/wp-content/uploads/2020/06/AppalachianDiseasesofDespairAugust2017.pdf.

Monnat, S. M. (2020). Working-age non-Hispanic white mortality: Rural–urban and within-rural differences. *Population Research and Policy Review*, **39**, 805–834.

Otani, T. and Takahashi, K. (2021). Flexible scan statistics for detecting spatial disease clusters: The rflexscan R package. *Journal of Statistical Software*, **99**, 1–29.

PARC (1964). Appalachia: A report by the president's Appalachian commision. https://www.govinfo.gov/content/pkg/GOVPU-Y3_AP4_2-PURL-LPS99948/pdf/GOVPUB-Y3_AP4_2-PURL-LPS99948.pdf.

Pollard, K. and Jacobsen, L. A. (2021). The Appalachian region: A data overview from the 2015-2019 American community survey chartbook. The Appalachian Regional Commission. https://www.arc.gov/report/the-appalachian-region-a-data-overview-from-the-2015-2019-american-community-survey/.

Quinones, S. (2015). *Dreamland: The True Tale of America's Opiate Epidemic*. Bloomsbury Press.

Rao, A. S., Pyne, S., and Rao, C. R. (2017a). *Handbook of Statistics: Disease Modelling and Public Health, Part A*, volume 36. Elsevier.

Rao, A. S., Pyne, S., and Rao, C. R. (2017b). *Handbook of Statistics: Disease Modelling and Public Health, Part B*, volume 37. Elsevier.

Ray, S., Desai, M., and Pyne, S. (2022). Identifying patterns of association in the use of addictive substances over five decades in the United States. *Computers in Biology and Medicine*, **151**, 106175.

Rigg, K., Monnat, S. M., and Chavez, M. N. (2018). Opioid-related mortality in rural America: Geographic heterogeneity and intervention strategies. *International Journal of Drug Policy*, **57**, 119–129.

Shiels, M. S., Tatalovich, Z., Chen, Y., and et al (2020). Trends in mortality from drug poisonings, suicide, and alcohol-induced deaths in the United States from 2000 to 2017. *JAMA Network Open*, **3**, e2016217.

Spillane, S., Shiels, M. S., Best, A. F., Haozous, E. A., and et al (2020). Trends in alcohol related deaths in the United States, 2000-2016. *JAMA Network Open*, **3**, e1921451.

Stacy, S., Chandra, H., Guha, S., Gurewitsch, R., Brink, L., Robertson, L., Wilson, D., Yuan, J.-M., and Pyne, S. (2023). Re-scaling and small area estimation of behavioral risk survey guided by social vulnerability data. *BMC Public Health*, **23**, 184.

Tango, T. and Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*, **4**, 1–15.

NACo and ARC (2019). Opioids in Appalachia. the role of counties in reversing a regional epidemic. https://www.naco.org/sites/default/files/documents/opioids-full.pdf.

U.S. Department of Health and Human Services (2023). What is the opioid epidemic? https://www.hhs.gov/opioids/about-the-epidemic/index.html.

Wallace, M., Sharfstein, J. M., Kaminsky, J., and Lessler, J. (2019). Comparison of U.S. county-level public health performance rankings with county cluster and national rankings: Assessment based on prevalence rates of smoking and obesity and motor vehicle crash death rates. *JAMA Network Open*, **2**, e186816.

Woolf, S. H., Chapman, D. A., M, B. J., Bobby, K. J., Zimmerman, E. B., and Blackburn, S. M. (2018). Changes in midlife death rates across racial and ethnic groups in the United States: systematic analysis of vital statistics. *BMJ*, **362**, k3096.

Woolf, S. H., Schoomaker, H., Hill, L., and Orndahl, C. M. (2019). The social determinants of health and the decline in U.S. life expectancy: implications for Appalachia. *Journal of Appalachian Health*, **1**, 6–14.