Original Research

The Journal of Prevention of Alzheimer's Disease - JPAD©

Geographic Clusters of Alzheimer's Disease Mortality Rates in the USA: 2008-2012

R.W. Amin[1], E.M. Yacko[1], R.P. Guttmann[2]

1. University of West Florida, Departments of Mathematics and Statistics, 11000 University Pkwy. Pensacola, Florida. 32514. USA, 2; University of West Florida, Departments of Biology. 11000 University Pkwy. Pensacola, Florida. 32514. USA

Corresponding Author: Rodney P. Guttmann, 11000 University Pkwy. Pensacola, Florida 32514. USA rpg@uwf.edu, 850-346-1414 (phone), 850-346-3182 (fax)

J Prev Alz Dis 2018 inpress
Published online inpress

Abstract

IMPORTANCE: The results identified geographic clusters of high and low Alzheimer's disease (AD)-related mortality across the contiguous United States. These clusters identify specific geographic groupings of counties that allow researchers to narrow the focus to identify some of the biopsychosocial variables contributing to increased or decreased AD mortality.

OBJECTIVES: To determine the extent to which geographic clusters exist where AD mortality significantly differs from the national average. Such knowledge could further future research in a more focused study of variables that are contributing to these differences.

DESIGN: Age adjusted AD mortality rates were analyzed with a spatial cluster analysis using the disease surveillance software SatScanTM.

RESULTS: Three large clusters had elevated age-adjusted AD mortality of at least 60% above the national average. These clusters were in Washington State, Iowa, and North and South Dakota. Below average AD mortality was observed in several areas including New York City, and parts of Arizona, California, Arkansas and Texas.

CONCLUSION AND RELEVANCE: This study demonstrates the use of disease surveillance methodology in identifying geographic patterns of unusually high or low AD mortality rates in the USA. Such results provide supporting evidence of appropriate locations to test interventions with the goal to reduce AD mortality.

Key words: Mortality risk assessment, spatial analysis, epidemiology, disease surveillance, SatScanTM, modeling.

Introduction

The main objective for this study is the identification of geographic areas in the contiguous USA where age-adjusted AD mortality rates are significantly higher (or lower) than the rest of the country.

AD mortality is underreported

According to a National Center for Health and Statistics data brief using data for the year 2010, (1). age adjusted Alzheimer's disease (AD) mortality rates were on the rise for the years 2000-2010, while other major diseases such as stroke, heart disease, and cancer were decreasing. It is expected that with no treatment or cure, there will be a continual rise in AD mortality. While AD is the sixth leading cause of death based on death certificates, these numbers do not fully gauge the disease's prevalence (2). As the cases of AD continues to rise, knowledge of any geographical aspects that may play a role in AD mortality can help guide researchers, advocates and government agencies utilize their resources more strategically.

Genetic variation in AD risk

While age is the greatest risk factor for AD, genetics also play an important role. Familial AD, those with a genetic cause, represents about 5% of cases and typically occurs in those younger than 65 (3). Late-onset AD represents about 95% of cases and may or may not have an associated genetic risk. The major genetic risk associated with late-onset is the presence of APOE-ε4 allele (4). When present, APOE-ε4 not only increases AD risk (5) but may also influence responses to drugs in various clinical trials targeting amyloid (6).

Psychosocial variables as contributors to AD mortality

AD mortality varies based on genetic complexity and the presence of additional regional biopsychosocial variables. For example, those with heart or heart related diseases are at an increased risk for AD (7). Additionally, smoking, obesity, high cholesterol, high blood pressure, excessive alcohol drinking, head trauma, depression, and low levels of formal education are also under investigation for effects on the rate of dementia (7-12). Difference in environmental chemical exposures are another potential geographic factor that could impact mortality and AD. Psychosocial variables including aspects of quality dementia care, access to general healthcare, cultural issues with stigma associated with mental health among others are all potential contributors to a community's risk of AD mortality. Using the current

Received April 30, 2018
Accepted for publication May 22, 2018

approach to identity areas of elevated/decreased AD mortality should allow for further, focused inspection of the causes that could include behavioral, health, genetic, or environmental factors.

Methods

Data Sources

Cumulative count data for Alzheimer's disease mortality for all counties in the contiguous United States were obtained for all years 2008-2012 from the CDC Wonder Multiple Cause of Death Database (13). Of the 3,109 counties, the data were suppressed by the CDC in 309 counties because the data did not meet the minimum privacy requirement of "at least ten AD deaths for a particular county". These 309 suppressed counties' cumulative death counts were estimated. Missing county values were replaced by the average of the state's AD mortality rate, multiplied by the county's population. This was done to minimize the effect of missing data on the cluster analysis performed. The CDC's AD data was matched with geographic information for each county with ArcGIS software (Redlands, CA) utilizing a 2010 TIGER/Line Shapefile (14) and using each county's identification number.

Disease Surveillance Software

In this study, we used the disease surveillance software SaTScanTM (15) to identify and test for the significance of AD mortality clusters. The mortality counts in each cluster were used in a two-dimensional spatial analysis for spatial effects. We assumed that the number of AD mortality cases in each county were distributed according to a Poisson model. This method tested the null hypothesis that the risk of AD mortality is the same for all counties in the continental United States. Basically, the expected number of AD mortality cases is proportional to each county's population size under the null hypothesis. In addition to AD case numbers and matching population counts for each county, the geographical coordinates are required for each county. This cluster analysis approach assigns a score based on a likelihood ratio statistic to each potential AD cluster under consideration to measure how unusual the AD rate observed inside the potential cluster is, as compared to the AD rate outside the potential cluster. Then, it evaluates whether the highest scores are unusually extreme compared to the distribution of high AD scores we would expect under the null.

The spatial scan statistic in SaTScanTM identifies clusters by imposing a window that moves over a map, including different sets of neighboring counties represented by their corresponding centroids. If the window includes the centroid of a specific county, then this county is included in the window. The center of the window can only be positioned at county centroids. For each window, the spatial scan statistic tests the null hypothesis of equal risk of AD mortality for all counties against the alternative hypothesis that there exists an elevated risk. Similarly, SatScanTM can be used for a low scan to identify geographic regions with reduced risk of AD mortality.

Statistical Modeling

The likelihood function for the Poisson model can be shown to be proportional to where n is the number of AD death counts within the scan window, N is the total number of AD deaths in the population, and E is the expected AD death count under the null hypothesis. Since we were using a one-tailed test that rejects the null hypothesis if there exists elevated AD risk, an indicator function I was used. The same methodology was used with scans for low AD mortality rates. If the research hypothesis was true, the indicator function set I equal to one when the scan window had a larger number of AD mortalities than expected; otherwise, it set it to zero. For a given N and E, it could be shown that the likelihood increased as the number of AD deaths, n, increased in the scan window. In this study, we present AD mortality clusters identified by circular windows with SaTScanTM.

Results

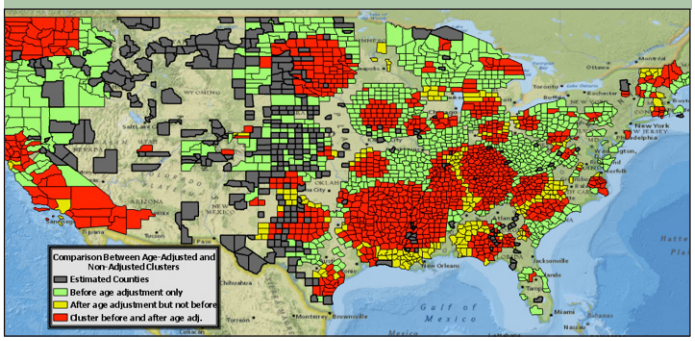
Spatial analysis of US

A purely spatial analysis was done in SaTScanTM using the Poisson model applied to Alzheimer's death counts per county for each of the states in the continental United States. The initial scan of the raw AD mortality rate resulted in 63 statistically significant clusters with p-value < 0.01 (data not shown). These clusters showed areas of significant elevated Alzheimer's mortality for the years 2008-2012.Due to the general effects of age on mortality, Alzheimer's mortality counts for each county were adjusted based on the age distribution of the citizens in each county. This was done by adjusting the data using five year age groups obtained from the United States Census Bureau age data and estimates for the years 2008-2012 (16).

In Figure 1a it was observed that age-adjusting resulted in changes in the clustering of counties, as was expected since AD rates are higher among older populations. The counties in red show elevated AD mortality that is not associated with age. Such clusters are not explained by increased age alone. Counties in green are cluster counties that are associated with age. In such cluster counties increases in AD mortality rate is associated with increased age within the county. Yellow colored counties indicate a county that appears not to have increased AD mortality, until age-adjustment

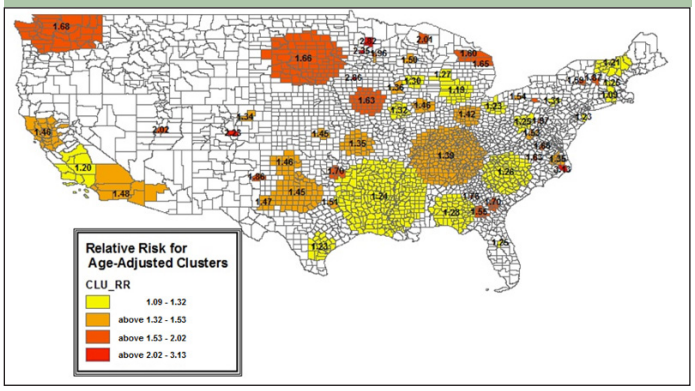
is made. This implies that for the age distribution of (yellow) counties, the matching AD rate is too high (or too low). Figure 1b shows the relative risk of the counties after age-adjustment. While Figure 1a identified the significant AD clusters, this map gives the corresponding Relative Risk (RR) value for each cluster. A RR of 1 for a cluster implies that the AD rate in this cluster is equal to the average AD rate in the contiguous USA. These data suggest that there is a non-random distribution of AD mortality across the United States. The RR values range from 1.09-132 for the yellow colored clusters, all the way to 2.02-3.13 for the red clusters, where the AD rate is between 2.02 to 3.43 times as high as the rest of the population.

Figure 1a. Clusters of AD mortality before and after adjusting for age as a covariate



Light green clusters represent counties with elevated AD mortality that are associated with increased age. Yellow counties are counties that were only identified with having high AD mortality after adjusting for age. Red counties have high AD mortality that is not associated with increased age. Counties in gray are those counties suppressed by the CDC because the data did not meet the minimum privacy requirement of "at least ten AD deaths for a particular county".

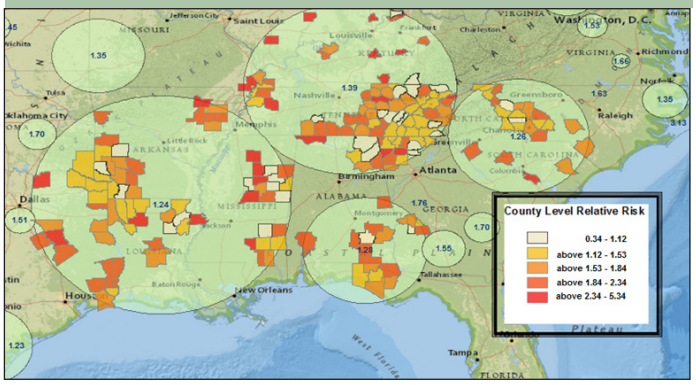
Figure 1b. Relative risk of clusters remaining after adjusting for age



Cluster map for the remaining clusters after age- adjustment and their relative risk denoted by the number in their center.

While large sized "windows" identified major clusters four large clusters in the southeast United States were further analyzed using SaTScanTM to refine areas in which there was a significant difference in age-adjusted AD mortality rates between groups of counties inside and the cluster as a whole (Figure 2). Each of the 4 clusters was re-analyzed by itself with SatScanTM to identify the hotspots within each cluster. In this analysis, the counties within each cluster become the "total population" for the spatial analysis. The counties in red show the highest risk of these areas with over twice the mortality rate (RR > 2.34) compared to the rest of the country. The largest region contained in these south eastern United States clusters is surrounding the city of Charlotte with population in five counties of approximately 1.5 million and mortality rates 61% higher than the rest of the country (RR= 1.61). Figure 2 shows that a "high AD cluster" may contain counties with high AD mortality, in addition to counties with lower AD mortality. The probability model takes this fact into account, and SatScanTM uses the average risk of AD within each identified cluster.

Figure 2. Cluster map detail of counties that belonged to clusters that had a significant difference when compared to their whole cluster



The counties are shaded based on their relative risk in reference to the whole United States. Circles outside these county shaded clusters represent other clusters that were not further examined and show their corresponding relative risk.

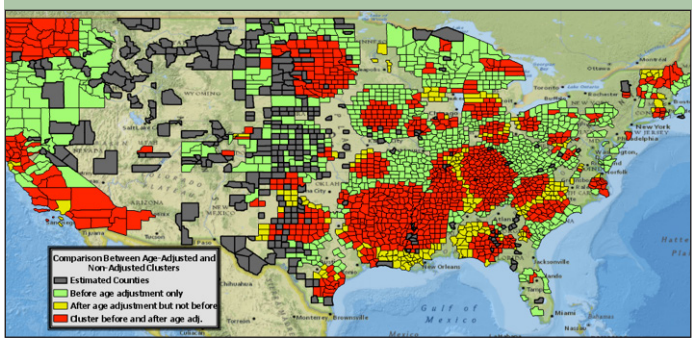
Next, we considered geographic clusters of low AD mortality. Figure 3 identifies clusters with AD mortality rates with relative risk less than 1.0. Note that some of these low AD mortality rate clusters may include counties that were included in some of the less pronounced high-risk clusters. This is possible because not all counties in a cluster necessarily possess high risk but may be included because they are surrounded by other counties that do.

Similar to what was observed for high risk clusters, the low risk clusters are spread throughout the United States. Interestingly, New York City and its surrounding counties had the lowest risk of all areas in which there was more than 100,000 in population. New York City had relative risk of less than half the average AD mortality rate (relative risk of 0.42). The largest cluster observed covers over 100 counties and parts of eight states including the large cities of Las Vegas and Salt Lake City with AD mortality rate 43% less than the average.

After assessing the locations across the United States with elevated mortality risk and those with low risk, we next obtained information on demographic factors that

could be associated with increased AD-related mortality (13, 16-26).

Figure 3. Low AD mortality risk clusters (P value < .01)



Cluster map showing the location of low AD mortality rates and the relative risk ratios.

The most significant positive / negative pairwise correlations between factors and age-adjusted Alzheimer's mortality are given in Table 1. Several of these factors had a significant pairwise correlation with AD mortality. The strongest pairwise associations with age-adjusted mortality are cancer mortality, physical inactivity, obesity, and heart disease. Diabetes, cancer mortality, percent less than HS diploma, and cancer air pollutants all increased in strength in their association after age-adjustment. Glyphosates, however, decreased in strength.

Table 1. Covariates showing strong associations with age-adjusted AD mortality ratio

	Pearson Correlation Analysis			
	Non-Adjusted AD	p-value	Age Adjusted AD	p-value
AD Rate before Age-Adj.	1		0.81027	<.0001
AD Age Adjusted Ratio	0.81027	<.0001	1	
Obesity Percent Age Adj.	0.1986	<.0001	0.21745	<.0001
Smoking Rate	0.21654	<.0001	0.20676	<.0001
% Less than HS Diploma	0.04227	0.0184	0.14648	<.0001
% with Bachelor's Degree	-0.17348	<.0001	-0.14091	<.0001
Diabetes Percent	0.04227	0.0184	0.19466	<.0001
Physically Inactive Percent	0.20589	<.0001	0.22729	<.0001
Heart Disease Mortality	0.20773	<.0001	0.21915	<.0001
Cancer Mortality	0.1777	<.0001	0.22866	<.0001
Glyphosates Pesticide	0.2255	<.0001	0.13796	<.0001
Cancer Air Pollutants	-0.14277	<.0001	0.15061	<.0001

Discussion

Rationale and context for study

Although it is still not known which variables are causal for increased mortality risk in those with AD, we sought to utilize this modern epidemiological surveillance approach to identify geographic areas with exceptionally high or low mortality due to AD. For this study, the term "environment" refers to all aspects of the local biopsychosocial environment that could include genetic predisposition, cultural caregiving norms, healthcare quality/access, environmental hazards, public policy effects, migration impact or yet unknown variables associated with a geographic region.

Previous work by Bagheri et al. (27) provides an extensive review of the current literature on geospatial analysis techniques for mapping dementia risk. The GIS publications on Dementia data were categorized by authors, year, risk predictors, and by risk profile. In their study, it is pointed out that most research on dementia used population data retrospectively. It is concluded that the application of GIS based methodologies to identify Dementia hotspots can provide information on possible geographical variations in risk factors. Bagheri (27) did not find any publication on a cluster analysis of Dementia for the USA. In other work, Bagheri et al. (28) used general practice clinical data for a cluster analysis of AD data for the city of West Adelaide (Australia) using spatial software algorithms based on Getis-Ord Gi* and Moran's I was also used. The authors used the Australian National University Alzheimer's Disease Risk Index to estimate the dementia risk. The study concluded that high dementia risk scores were associated with high age, high cholesterol, living alone, physical inactivity, diabetes, and depression. Cluster analyses can be based on different algorithms. While Bagheri (28) chose to use Getis-Ord Gi* and Moran's I, the present study used the disease surveillance software SatScanTM on AD mortality rates at the county level. To the best of our knowledge, the present study is the only study on AD mortality where the disease surveillance software SatScanTM is used for all counties in the contiguous USA.

AD mortality is not evenly distributed in the United States

As the ability to treat other causes of early mortality improve, those surviving past the age of 65 are facing the possibility of developing AD for which there is no treatment or cure. Unfortunately, there is the compounded issue of early mortality in those with AD due to other co-morbid conditions that may go undisclosed (9). Utilizing available county-level data, regions in US with exceptional mortality rates provides an opportunity to focus efforts in the most critical areas. Additionally, this study identified areas with low mortality that may represent locations with best practices to be studied and replicated across the country. With the limits of accurate data reporting, the disease surveillance approach used in this study identifies groups of counties for further research by health authorities and academic researchers to better understand AD mortality.

Limitations and Strengths

This work is limited by the availability of accurate reporting and recording of data. It is possible that there is incomplete data reporting from some of the counties. This should be mitigated by the process of reporting results at the county level to state level and then to the CDC. Another limitation is the possibility of inaccurate recording of AD as the cause of death (2). These phenomena could be more pronounced in some regions compared to others where the impact of regional Alzheimer's research centers or advocacy could exert influence on the documentation of cause of death. Another limitation of this study is the fact that certain geographic regions may be impacted by at-risk individuals migrating to specific areas and therefore biasing the analysis. Strengths of the study include (i) No preselection bias for any cluster (ii) Extensive coverage of US counties (iii) Adjusting AD mortality rates for age with regression analysis (iv) Clusters are tested for significance based on a likelihood ratio (v) Identified clusters assist researchers in further studies which incorporate covariates with potential roles in the unusually high (or low) AD mortality rates.

Future Directions

Using disease surveillance methodology opens the door for focused research in specific geographic areas. Additional studies in the clusters identified here should facilitate a more cost-effective approach by narrowing efforts to understand cause and effect of AD mortality where the data suggest need is greatest. Second, in those areas where there is low risk, identification of potential best practices is recommended. Alternatively, such differences measured here may indicate that rather than having better or worse health outcomes, reporting is not accurate. Audits of reporting that indicated such was the case, efforts to improve accurate reporting to the CDC would be appropriate. Inaccurate diagnosis or documentation of AD as a cause of death is another area warranting investigation as noted by others (2) that could be implemented in those areas with below average AD mortality rates to confirm the present results.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest: ?????

Ethical standards: ???????.

References

1. Tejada-Vera, B., [dataset] Mortality from Alzheimer's disease in the United States: Data for 2000 and 2010. NCHS data brief, 2013. 116.
2. Romero, J.P., et al., Under reporting of dementia deaths on death certificates: a systematic review of population-based cohort studies. J Alzheimers Dis, 2014. 41(1): p. 213-21.
3. Mendez, M.F., Early-Onset Alzheimer Disease. Neurol Clin, 2017. 35(2): p. 263-281.
4. Profenno, L.A., A.P. Porsteinsson, and S.V. Faraone, Meta-analysis of Alzheimer's disease risk with obesity, diabetes, and related disorders. Biol Psychiatry, 2010. 67(6): p. 505-12.
5. Farrer, L.A., et al., Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. Jama, 1997. 278(16): p. 1349-56.
6. Petersen, R.C., et al., Randomized controlled trials in mild cognitive impairment: Sources of variability. Neurology, 2017. 88(18): p. 1751-1758.
7. Snyder, H.M., et al., Vascular contributions to cognitive impairment and dementia including Alzheimer's disease. Alzheimers Dement, 2015. 11(6): p. 710-7.
8. Graves, A.B., et al., Alcohol and tobacco consumption as risk factors for Alzheimer's disease: a collaborative re-analysis of case-control studies. EURODEM Risk Factors Research Group. Int J Epidemiol, 1991. 20 Suppl 2: p. S48-57.
9. Heun, R., et al., Alzheimer's disease and co-morbidity: increased prevalence and possible risk factors of excess mortality in a naturalistic 7-year follow-up. Eur Psychiatry, 2013. 28(1): p. 40-8.
10. Janson, J., et al., Increased risk of type 2 diabetes in Alzheimer disease. Diabetes, 2004. 53(2): p. 474-81.
11. Ott, A., et al., Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study. Bmj, 1995. 310(6985): p. 970-3.
12. Ott, A., et al., Smoking and risk of dementia and Alzheimer's disease in a population-based cohort study: the Rotterdam Study. Lancet, 1998. 351(9119): p. 1840-3.
13. [dataset] Underlying Cause of Death 1999-2014 on CDC WONDER Online Database, released 2016. Data are from the Multiple Cause of Death Files, 1999-2015, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Centers for Disease Control and Prevention, National Center for Health Statistics; Available from: wonder.cdc.gov/ucd-icd10.html.
14. Branch, G.P., TIGER/Line® Shapefiles and TIGER/Line® Files. U.S. Census Bureau, 2012.
15. Kulldorff, M., SaTScanTM v9.4: Software for the spatial and space-time scan statistics. 2016.
16. Age and Sex, 2008-2012 American Community Survey 5-Year Estimates. U.S. Census Bureau.
17. Dwyer-Lindgren, L.M., A.H.; Srebotnjak, T; Flaxman, A.D.; Hansen, G.M.; Mruuar, C.J. , [dataset] Cigarette smoking prevalence in US counties: 1996-2012. Population Health Metrics, 2014.
18. [dataset] ACS DEMOGRAPHIC AND HOUSING ESTIMATES , 2008-2012 American Community Survey 5-Year Estimates U.S. Census Bureau, 2008-2012 American Community Survey, 2012.
19. [dataset] POVERTY STATUS IN THE PAST 12 MONTHS, 2008-2012 American Community Survey 5-Year Estimates., in U.S. Census Bureau, 2008-2012 American Community Survey. 2012.
20. [dataset] EDUCATIONAL ATTAINMENT, 2009-2013 American Community Survey 5-Year Estimates. U.S. Census Bureau, 2009-2013 American Community Survey, 2013.
21. [dataset] OCCUPANCY CHARACTERISTICS , 2008-2012 American Community Survey 5-Year Estimates. U.S. Census Bureau, 2008-2012 American Community Survey, 2012.
22. [dataset] 2011 NATA Emissions by County. United States Environmental Protection Agency, 2015.
23. [dataset] Baker, N.T., and Stone, W.W, Estimated annual agricultural pesticide use for counties of the conterminous United States, 2008–12. U.S. Geological Survey Data 2015. 907: p. 9.
24. [dataset] Daily Fine Particulate Matter (PM2.5) ($\mu g/m^3$) for years 2003-2011. CDC WONDER Online Database, 2012.
25. [dataset] Diagnosed Diabetes Prevalence, Obesity Prevalence, Leisure-time Physical Inactivity Prevalence. Centers for Disease Control and Prevention 2016.
26. [dataset] Diagnosed Diabetes Prevalence Centers for Disease Control and Prevention, 2016.
27. Bagheri, N., et al., Combining Geospatial Analysis with Dementia Risk Utilising General Practice Data: A Systematic Review. J Prev Alzheimers Dis, 2018. 5(1): p. 71-77.
28. Bagheri, N., et al., General Practice Clinical Data Help Identify Dementia Hotspots: A Novel Geospatial Analysis Approach. J Alzheimers Dis, 2018. 61(1): p. 125-134.