Research

Evaluating the disparity of female breast cancer mortality among racial groups - a spatiotemporal analysis

Chiehwen Ed Hsu*[1,2], Holly Jacobson[3] and Francisco Soto Mas[3]

Address: [1]Department of Health Management and Policy, School of Public Health, University of North Texas Health Science Center, 3500 Camp Bowie Blvd, ME1-740. Fort Worth, TX 76107. USA, [2]School of Health Information Sciences, University of Texas-Houston Health Science Center, 7000 Fannin, UCT 600, Houston, TX 77030. USA and [3]Department of Social and Behavioural Sciences, School of Public Health, University of North Texas Health Science Center. 3500 Camp Bowie Blvd, ME1-711. Fort Worth, TX 76107. USA

Email: Chiehwen Ed Hsu* - chhsu@hsc.unt.edu; Holly Jacobson - hjacobso@hsc.unt.edu; Francisco Soto Mas - fsotomas@hsc.unt.edu

* Corresponding author

Abstract

Background: The literature suggests that the distribution of female breast cancer mortality demonstrates spatial concentration. There remains a lack of studies on how the mortality burden may impact racial groups across space and over time. The present study evaluated the geographic variations in breast cancer mortality in Texas females according to three predominant racial groups (non-Hispanic White, Black, and Hispanic females) over a twelve-year period. It sought to clarify whether the spatiotemporal trend might place an uneven burden on particular racial groups, and whether the excess trend has persisted into the current decade.

Methods: The Spatial Scan Statistic was employed to examine the geographic excess of breast cancer mortality by race in Texas counties between 1990 and 2001. The statistic was conducted with a scan window of a maximum of 90% of the study period and a spatial cluster size of 50% of the population at risk. The next scan was conducted with a purely spatial option to verify whether the excess mortality persisted further. Spatial queries were performed to locate the regions of excess mortality affecting multiple racial groups.

Results: The first scan identified 4 regions with breast cancer mortality excess in both non-Hispanic White and Hispanic female populations. The most likely excess mortality with a relative risk of 1.12 (p = 0.001) occurred between 1990 and 1996 for non-Hispanic Whites, including 42 Texas counties along Gulf Coast and Central Texas. For Hispanics, West Texas with a relative risk of 1.18 was the most probable region of excess mortality (p = 0.001). Results of the second scan were identical to the first. This suggested that the excess mortality might not persist to the present decade. Spatial queries found that 3 counties in Southeast and 9 counties in Central Texas had excess mortality involving multiple racial groups.

Conclusion: Spatiotemporal variations in breast cancer mortality affected racial groups at varying levels. There was neither evidence of hot-spot clusters nor persistent spatiotemporal trends of excess mortality into the present decade. Non-Hispanic Whites in the Gulf Coast and Hispanics in West Texas carried the highest burden of mortality, as evidenced by spatial concentration and temporal persistence.

Background

According to a recently released U.S. Cancer Statistics report, breast cancer was the leading cause of cancer deaths among American women in all racial/ethnic groups in the year 2000. [1] National and state-specific studies indicated that the distribution of breast cancer mortality varied by race in Texas over an extended period of time. Spatiotemporal variations observed by the Cancer Atlas of the National Cancer Institute [2] suggested that the Houston-Galveston and Dallas-Fort Worth State Economic Areas (SEAs) had the highest breast cancer mortality among White females. Among the Black female population, Abilene and Alice SEAs had the highest breast cancer mortality between 1970 and 1994. Texas-specific cancer research also indicated that Texas counties near the Gulf Coast, Bexar and El Paso counties, among others, had an excess mortality from cancers between 1980 and 1997. [3-5] On the other hand, the report of U.S. Cancer Statistics suggested that disparities exist in cancer mortality among different racial groups. [1] The report indicated that the occurrence of breast cancer among non-Hispanic White women was almost 1.2 times higher than that of Black women, and 1.7 times higher than that among Asians/Pacific Islanders. In Texas, 26,338 females died of cancer in the last decade. Among them, 72% (18,966) were non-Hispanic White females.[6] Conversely, other research argued that excessive breast cancer mortality presented an uneven burden on African-Americans, as this particular racial group experienced worse breast cancer outcomes, [7] and that African-American and Hispanic women had poorer overall survival rates from breast cancer.[8] Although the literature provides inconclusive results in terms of which race/ethnicity may suffer the most from the burden of breast cancer mortality, it nevertheless underlines the importance of clarifying the spatiotemporal disparity in racial groups.

To quantify the breast cancer mortality burden by race across space and time, this study adopted a statistical approach to characterize the spatiotemporal clusters of breast cancer mortality. A "cluster", in this context, is detected within a defined geographical area during a specific timeframe when the area has a disproportionate excess in mortality, when compared to the neighbouring areas under study.[9] By meeting the statistical assumptions of a set of statistical models, the unusual rise or reduction of mortality in a specific spatial and temporal window (with adjustments for demographic factors such as age and gender, or other substantiated risk factors) can be characterized by statistical significance. In this context, this study used the terms "clusters" and "excess mortality" interchangeably, with both terms referring to the statistical context of both spatial and temporal dimensions of excess.

For the time period between 1990 and 2001, the present study evaluated the county-level excess of breast cancer mortality in three predominant racial groups of Texas female populations. The excess mortality burden was characterized by spatiotemporal variations. The study tested the potential continuation of excess deaths for 10 years or more to the present decade. Based on the results of analysis, the study identified each racial group and multiple groups in Texas regions that may have been most affected by the persistent mortality burden over time. These results point to priority geographic areas for policy deliberation.

Data collection and treatments

To identify potential breast cancer excess mortality in Texas counties between 1990 and 2001, breast cancer race-specific deaths, the female population at risk, and location data were collected and saved in three separate files. The first, "Deaths Files By Race", included female breast cancer deaths (ICD-9 Code 174 and ICD-10 Code C50), which reported the place of residence in 254 Texas counties of 4 racial groups, that were coded as categorical data (e.g., non-Hispanic White = 1, Black = 2, Hispanic = 3 and Other = 4) in each of the 12 study years. Each file contained 16 age-group categorical variables, with values ranging from 1 to 16, representing the ages of "0 to 4" to "75 and above", grouped at 5-year intervals. The data were so arranged for age adjustment and race stratification. One file for per racial group was created, including 48,768 records reflecting the number of deaths for each race among the 16 age groups in 254 counties over the 12-year study period. The second file, the "Population File", contained data on the populations at risk in the study period, (i.e., the female population in each Texas county, during the 12 study years, with respect to race and age group information that corresponded to the "Deaths File By Race"). Race and age data from the Year 2000/1990 Census were obtained from the "Summary File 1" of Census 2000/1990, originated from the American FactFinder Website of the U.S. Census Bureau. [10] The population data for the remaining years were obtained from population estimates made available through the Texas State Data Center and the Center of Vital Statistics of the Texas Department of Health. [11] The Population file contained a total of 195,072 records, representing the four races and 16 age groups in 254 counties for the 12-year study period. Age and race variables were also coded as categorical data to enable subsequent adjustment and stratification. The third file, the "Geographic File", was also obtained from the US Census Bureau. [12] This file contained the latitude and longitude information of Texas county centroids as a proxy that indicated the locality of each county. The Texas county shapefiles were obtained from the CDC Website (URL http://www.cdc.gov/epiinfo/usa/tx.exe) for further mapping analysis.

Methods of analysis

This study employed the Spatial Scan Statistic developed by Kulldorff and colleagues [13,14] to detect potential excess breast cancer mortality. The test statistic was adopted previously for detecting excesses of breast and brain cancers. [14-16] When compared with other statistical methods for cluster detection, this statistic was found to have good power for detecting localized hot-spots type of excess events, particular those in rural areas.[17,18] This test statistic seems appropriate for detecting potential excess breast cancer mortality in the unique rural-urban combination of the state of Texas.

The Spatial Scan Statistic factors in uneven geographical population densities and conditions, and then analyzes the total number of observed breast cancer deaths. It searches for clusters of cases without specif ying their size or location ahead of time, which tests for their statistical significance while adjusting for the multiple testing inherent in such a procedure. For an analysis of rare cases/deaths, such as cancer, the Poisson model can be used for estimating the probability distribution when the number of cases/deaths is substantially smaller than that of the population at risk. When there are no covariates, the null hypothesis of the Poisson model provides that the expected death counts in each county are proportional to the population size (or person-years) in that area. The alternative hypothesis states that deaths are not randomly distributed. In the present study, for each location and size of the scanned space and time, the alternative hypothesis refers to elevated adjusted mortality rates within space and time as compared to outside areas under study. Calculations can be performed using the SatScan Program (version 4.0, freeware available from URL:http://www.sat scan.org). SatScan first aggregates data with the scanning window of spatial (referring to the population at risk) – as a cylinder base, and temporal (years in this study) – corresponding to the height of the cylinder as selected by the users. For each cylinder, the scan adjusts for covariates and calculates the Log Likelihood Ratio (LLR, formula described below) by scanning through and plotting circles around geographic identifiers (prepared in the Geographic File) in a population size specified by the user across the entire study area. The base is the same as defined in spatial statistics, while the height reflects the time period of potential clusters. The cylindrical window moves in space and time and scans through each possible geographic location, defined by county centroids in the present study. The overall relative risk for each cluster, along with a set of simulated values based on the same procedure within a specific space and time, are then calculated. The latter are used as a baseline against the LLR values of the observed values. SatScan employs the Monte-Carlo simulation to estimate the LLR. When the LLR values of observed windows are higher than LLR based on simulation, SatScan determines the deaths in a particular region that are significantly different from the rest of the study area for the particular time window by rejecting the null hypothesis. Under the Poisson assumption, the Likelihood function for a specific space-time is then proportional to:

$$LLR = (c/n)^c([C-c]/[C-n])^{(C-c)}I() \quad [9]$$

Where C is the total number of breast cancer deaths, c is the number of cases within the space-time window, and n is the covariate(s)-adjusted expected number of deaths within the space-time analysis under the null-hypothesis. I() is an indicator function, whereby I() is equal to 1 when the timeframe has more deaths than is expected under the null-hypothesis, and is 0, otherwise. Based on a test statistic value of the LLR, a p-value is then calculated which suggests how well all the variables fit into the model at the same time. SatScan performs adjustments by indirect standardization.

For the present study, the Poisson model was used to calculate the number of expected deaths in each county. The space-time retrospective analysis was conducted without prior assumptions as to the size or location of such areas or duration of excessive mortality. The scan setting was set at a maximum spatial cluster size of 90% of the study period (i.e., 10 years) and 50% of the population at risk. The "50% of population at risk" parameter was recommended by Kulldorff [9] as an optimal value setting that maximizes the effect of potential cluster detection. This means that a cluster would comprise, at most, 50% of the population at risk. The study further tested the potential persistence of temporal clusters across the entire study period (i.e., 12 years) by holding constant the 50% maximum spatial cluster and scanning with the "purely spatial" option.

For data processing, we developed a Visual Basic application to automate data collection and manipulation, and output the results to geographic information systems (GIS) for performing mapping and spatial queries. The SatScan program saved the output files, including cluster locations, relative risk for each location, simulated LLRs, and the test statistic, in database (.dbf format) files. Data were stored on a Microsoft SQL Server version 7.0, and The SatScan program calculated LLR by performing 999 instances of Monte Carlo replications. The automated process, including data input, scanning, and output, took an average of 25 minutes of computer time.

Results

The study included 28,813 breast cancer deaths among an average female population of 9,585,195 in Texas counties across the 12-year study period. The age-and-race-
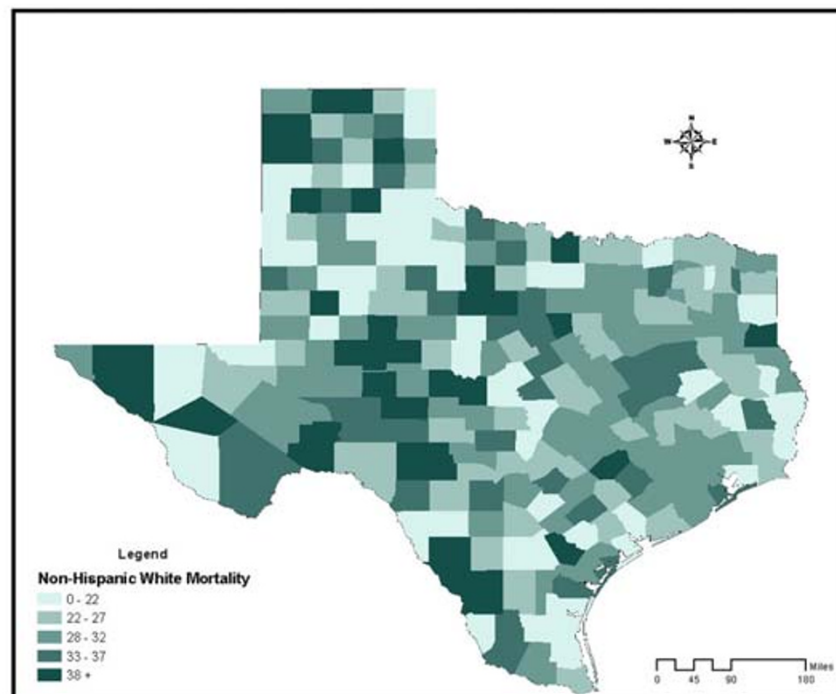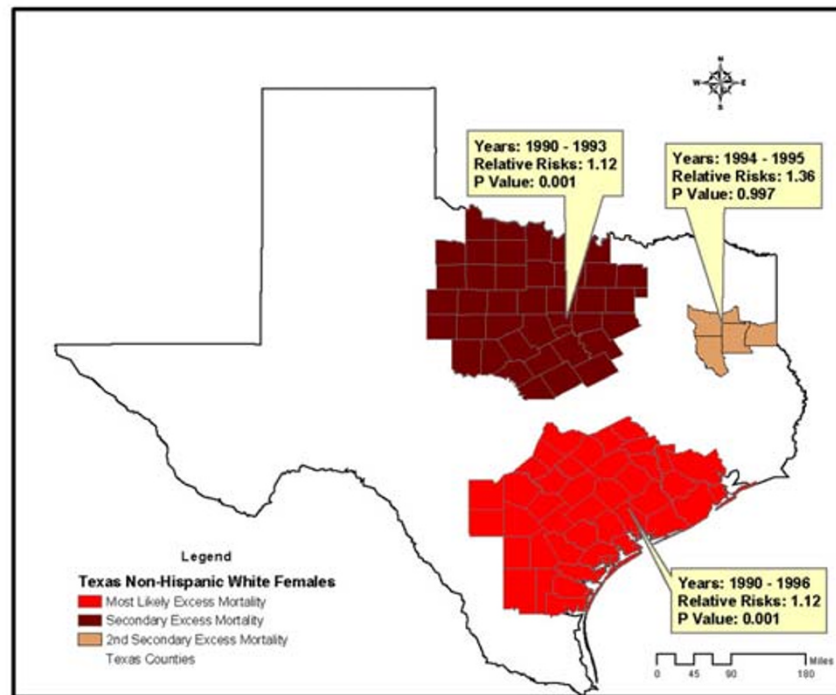
Figure 1
Potential excess of breast cancer mortality detected in the <u>non-Hispanic White </u>female population of Texas counties, 1990–2001, conducted with a scan window of a maximum of 90% of the study period and a spatial cluster size of 50% of the population at risk. Age-adjusted breast cancer mortality rates of <u>non-Hispanic White</u> Females in Texas Counties, 1990–1998.
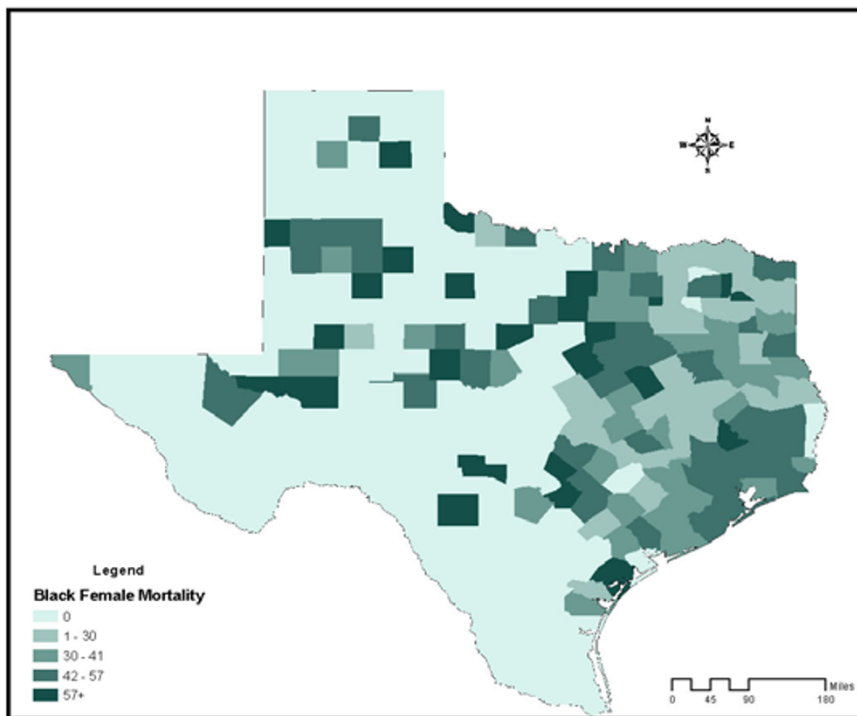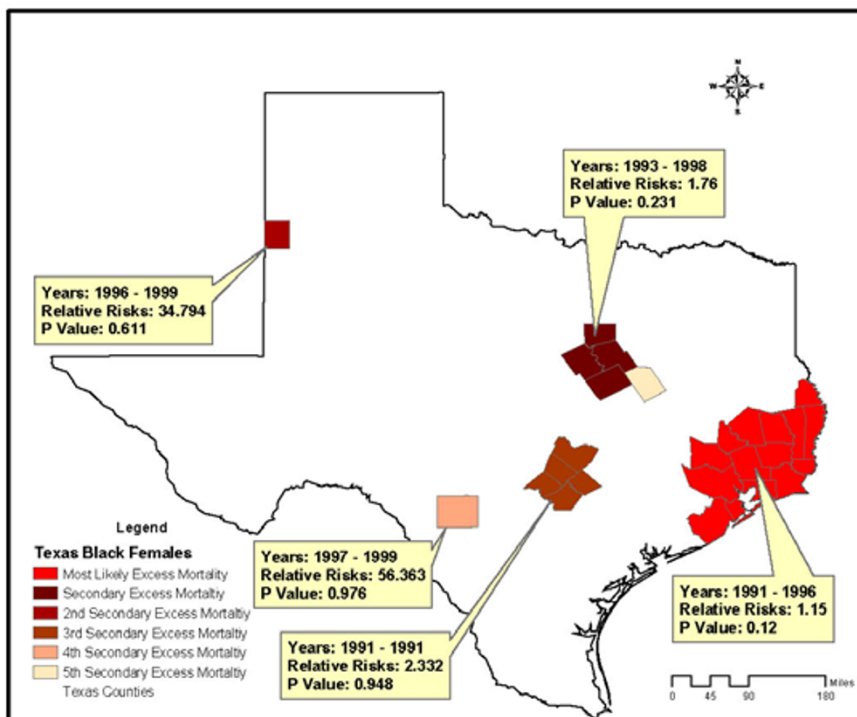
Figure 2
Potential excess of breast cancer mortality detected in the <u>Black</u> female population of Texas counties, 1990–2001, conducted with a scan window of a maximum of 90% of the study period and a spatial cluster size of 50% of the population at risk. Age-adjusted breast cancer mortality rates of <u>Black</u> females in Texas Counties, 1990–1998.
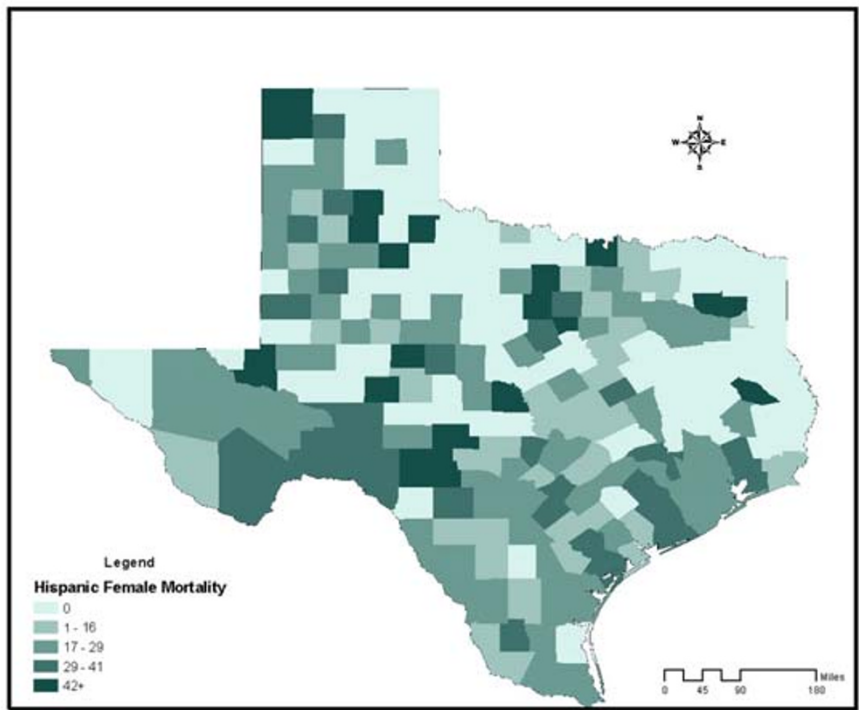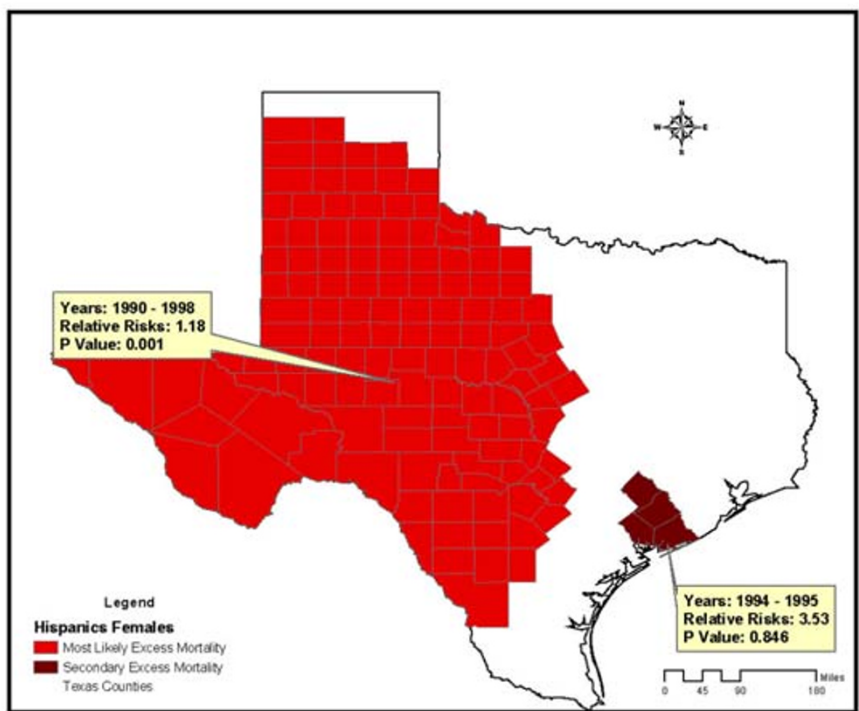
Figure 3
Potential Excess of Breast Cancer Mortality detected in the <u>Hispanic</u> female population of Texas Counties, 1990–2001, conducted with a scan window of a maximum of 90% of the study period and a spatial cluster size of 50% of the population at risk. Age-adjusted breast cancer mortality rates of <u>Hispanics</u> females in Texas Counties, 1990–1998.
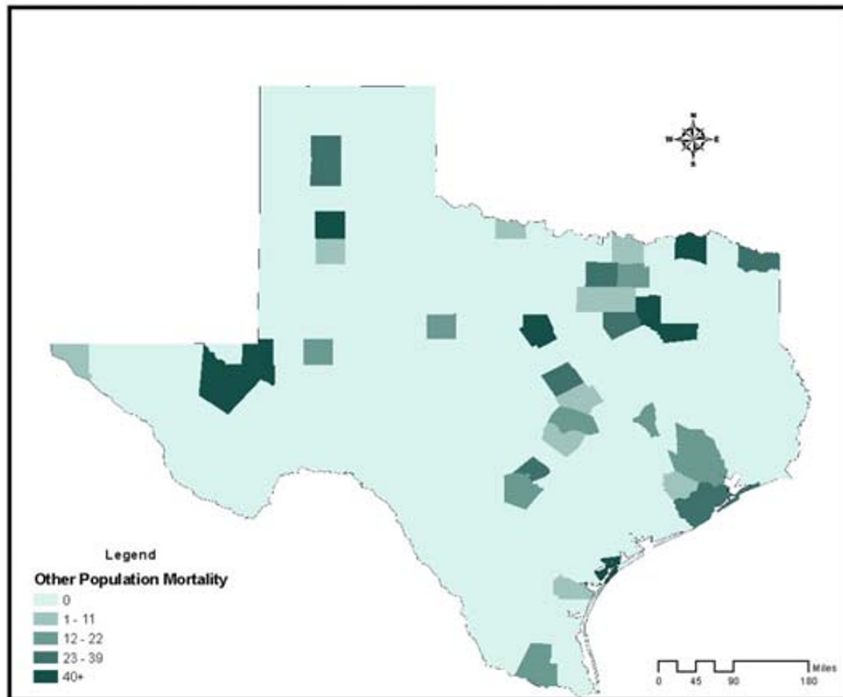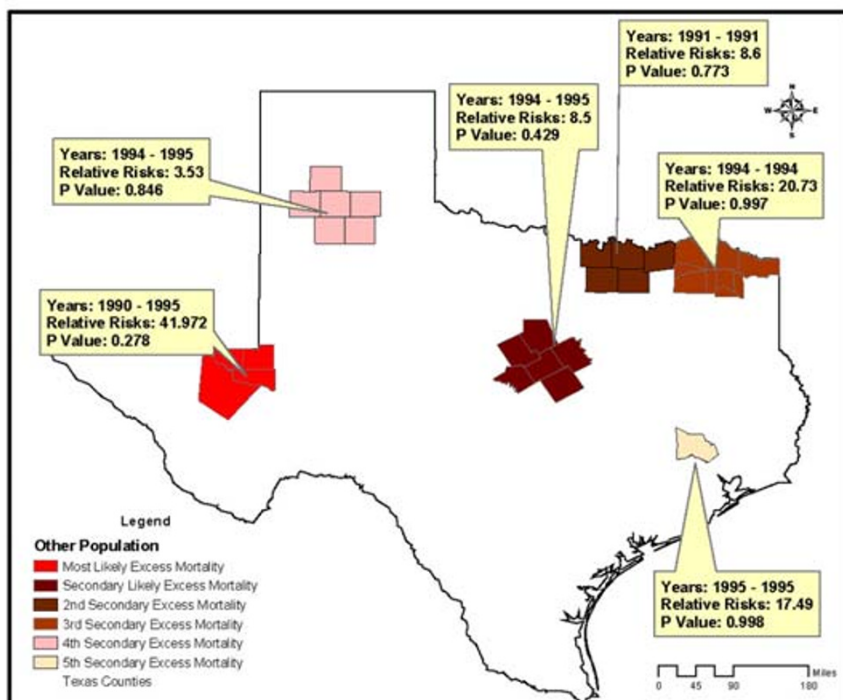
Figure 4
Potential Excess of Breast Cancer Mortality detected in the <u>Other</u> female population of Texas Counties, 1990–2001, conducted with a scan window of a maximum of 90% of the study period and a spatial cluster size of 50% of the population at risk. Age-adjusted breast cancer mortality rates of <u>Other</u> females in Texas Counties, 1990–1998.

Table 1: Texas counties detected with excess breast cancer mortality with statistical significance 1990–2001.

Potential Excess mortality	Annual Age-adjusted rates (per 100,000)	Year	Observed deaths	Expected deaths	Relative Risk (RR) of excess mortality	P Values	Total counties
Most Likely Cluster for non-Hispanic Whites (90% study period and 50%population at risk)	35.0	1990 – 1996	4088	3623	1.12	0.001	42
Secondary Cluster for non-Hispanic Whites	34.8	1990 – 1993	2240	1995	1.12	0.001	41
Primary Cluster for Hispanic White	29.0	1990 – 1998	1436	1212	1.18	0.001	127
Primary Cluster for Blacks (note: not statistically significant)	34.1	1991 – 1996	791	682	1.15	0.129	15

adjusted annual mortality rate of all races was 25/100,000 women/year. Annual age-adjusted mortality rates for non-Hispanic Whites, Blacks, and Hispanics were 31, 29.4, and 12.3 per 100,000 females respectively. Figures 2 to 4 are summary choropleth maps of the age-adjusted (to 2000 US population) mortality rates by quartiles at the Texas county level. With the adjustment of age and stratification by race, a scan window of a maximum of 90% of the study period (i.e., 10 years) and 50% of the population at risk revealed 10 regions of likely excess mortality in three racial groups within Texas female population. Among these, four regions were statistically significant in terms of both spatial and temporal excess. Figures 1 to 3 present these likely areas of excess mortality in this set of analysis. To describe the extensive geographic regions of the state of Texas in a consistent manner, we adopt the term used in the "counties and regions cross-reference" for each public health regions of Texas, as defined by the Texas Health and Human Services Commission. (URL http://www.hhsc.state.tx.us/about_hhsc/HHS_Regions.html). For the non-Hispanic White population, three potential excess mortality regions were detected. The most likely area of excess mortality with a Relative Risk (RR) of 1.12 (p = 0.001) occurred between 1990 and 1996 in southeast Texas along Gulf Coast, Central and Upper South Texas. These included 42 counties ranging from Harris to Kleberg counties. A secondary excess mortality region (RR = 1.12, p = 0.001) was identified between 1990 and 1993 in the 42 counties of Northwest and Metroplex Texas. One potential excess area in the non-Hispanic White population was detected in 5 counties in Upper East Texas that was not statistically significant (RR = 1.36, p = 0.99). For the Black population, the most likely area of excess mortality (RR = 1.15, p = 0.12) occurred between 1991 and 1996 in southeast Texas along the Gulf Coast. These included 15 counties ranging from Gasper to Brazoria. This cluster was not statistically significant. Additionally,

5 regions identified with potential excess mortality were not statistically significant (p > 0.20). For the Hispanic population, the most likely area of excess mortality (RR = 1.18, p = 0.001) occurred between 1990 and 1998 in West Texas, ranging from the border of the state of New Mexico to Lower South Texas. This area included 127 counties. An area of secondary excess mortality was found in the Gulf Coast and Upper South Texas (RR = 3.53, p = 0.848) between 1990 and 1995. For the "Other" population, six clusters were detected, and none of these were statistically significant (p > 0.28). It was not clear whether the temporal component of the clusters were there because it was truly a cluster restricted in time, or whether it is a purely spatial cluster that showed up as a space-time cluster due to the maximum temporal cluster size restriction. To further detect whether this cluster (or others) might have persisted into the current decade, a second scan analysis was performed with a "spatial only" option to enable the detection of 12-year clusters. The results of the second scan were almost identical to those of the first set of analysis, with the p value slightly increased in the most likely cluster for Blacks (p = 0.131). In terms of spatial persistence, the number of counties originally in the secondary possible excess mortality area remained the same. The secondary cluster remained the same in terms of relative risk and p-value. The non-statistically-significant cluster remained not statistically significant. Tables 1 summarizes the results of analysis by potential excess mortality, duration of occurrence, observed and expected deaths, counties, and relative risk information in each excess mortality region of a statistical significance.

To investigate those counties of excess mortality involving more than one racial group, we further conducted a spatial query. For both non-Hispanic Whites and Blacks, there were 3 counties in Gulf Coast Texas (Harris, Galveston and Brazoria counties) presented excess mortality,

and 9 counties in Central Texas (Hays, Comal, Guadalupe, Bexar, Medina, Wilson, Atascosa, Frio and McMullen) had excess mortality among non-Hispanic Whites and Hispanics.

Discussion

The results indicate that between 1990 and 1998, four geographic regions were identified with excess mortality rates in Texas that were statistically significant. With respect to suspected excess mortality, the regions detected with excess breast cancer mortality were consistent with those presented in the analyses of the 1970–1994 data by the US National Cancer Institute [2] and the 1990–1997 data reported by Zhan. [4] The results rendered supporting evidence that most counties that were previously suspected of having elevated breast cancer mortality do indeed have excessive cancer mortality. The present study additionally identified West Texas counties as having excess mortality from breast cancer in the Hispanic female population that persisted for 9 years, which was not previously reported. The relative risk of this cluster was at the modest level of 1.18. Nevertheless, this region had the highest in relative risk, with the longest temporal persistence among detected potential clusters of all racial groups in this study. Based on this finding and on the comparisons of LLRs for the primary suspected clusters from both scan trials (non-Hispanic Whites LLR = 35.00 vs. Blacks LLR = 10.01 vs. Hispanic LLR = 29.01 vs. Others LLR = 8.30), it was determined that the Hispanic and non-Hispanic White female populations in the regions detected with clusters had the highest burden of breast cancer mortality, as evidenced by both temporal persistence and spatial concentration.

The verification of breast cancer excess mortality over time may prove beneficial for health policy and planning. For instance, the state of Texas has yet to reach the Healthy People 2010 Objective of 16.3 deaths per 100,000 females in the population, as provided by the U.S. Centers for Disease Control and Prevention (i.e. Objective 3-3, "reduce the breast cancer death rate", http://www.healthypeople.gov/Document/html/tracking/od03.htm). Spatiotemporal analysis such as that described in this study will be instrumental in planning and reaching the projected objective. For instance, the present analysis underscored the two regions with multiple racial groups that bear the persistent burden of breast cancer mortality, and detected a potential 9-year persistence of excess breast cancer mortality with the highest relative risk in the Hispanic population. Both regions carry a disproportional burden of excess mortality and warrant further investigation and policy intervention. The results of spatiotemporal analysis quantified disease burden over time by both spatial concentration (as determined by p values, LLRs and relative risks) and temporal persistence (as determined by the

duration of detected clusters), which presented another perspective of measuring health disparity. It contributed to an understanding of the persistent burden of the disease across space and time, as well as aiding in determining whether the mortality burden that may have persisted into the current decade.

Several research notes arose from this study and warrant elaboration. First, as identified in the present study, the very modest relative risks that occurred over a large region of contiguous counties in Texas did not necessarily meet the strict definition of "clusters" of epidemic intensity. Compared with previous studies using SatScan for cluster detection, [16,19] the relative risks revealed in the present study were apparently lower, and no localized, hot-spot clusters (with constant, high risks in the clusters) that persisted over time were detected. On the other hand, breast cancer may have a substantial developmental period and may have potential risk/vulnerability factors, such as the stage at diagnosis, access to treatment, and the exposure to environmental toxic wastes that are not fully understood. These potential contributors were not accounted for in the present study. Given that many of these causes and risk factors may have operated over various time scales, the mortality examined here is only an endpoint in that process. While early detection of cancer is generally beneficial to survival, there is controversy over the effectiveness of breast cancer screening in reducing mortality.[20,21] Ideal interventions may also target modifiable risk factors that exist above and beyond the windows of space or time considered here. Nevertheless, this study offered baseline descriptions of persistently elevated breast cancer deaths in Texas, which may serve as a point of departure for policy deliberation and health resource allocation. Second, although this study focused primarily on statistically significant excess mortality, it by no means suggested that those non-statistically-significant regions of excess mortality were less important. To be statistically significant at the 0.05 or 0.01 level, outcome measures had to satisfy the Poisson distribution model and all independent variables of this study, including space, time and age, had to fit into the model simultaneously, and produce a large LLR as a result of spatial-temporal analysis. For example, the potential cluster detected among Blacks between 1991–1996 in Gulf Coast Texas (RR = 1.15, p = 0.12) was for all age groups. However, the results may become statistically significant if analysis was conducted with the stratification of certain age groups, such as among Black females aged 25 to 40. Therefore, the p-value derived is construed as an indicator, suggesting the level of excess mortality that calls for further investigation. Third, the choice of county level analysis entailed the strengths and weaknesses intrinsic to this level of aggregation. Although sub-county level (such as census tracts or block groups) of analysis may be preferred in cancer analysis, [22] we chose the county-level

data because this level of aggregation was used in other studies on detecting breast cancer clusters,[4,15] and also because the disproportional demographic distribution of Texas population made sub-county level analysis less feasible. For example, there were seven border counties that averaged fewer than ten breast cancer deaths, and had a population of less than 900 residents during the study period. The rates based on these small numbers of events and small population sizes tend to be unpredictable and often inflated. In particular, several of the above sparsely populated counties were in the summary choropleth maps of the age-adjusted mortality rates by race presented in Figures 1 to 3. These inflated mortality rates tend to produce visual bias, as these are the counties that attract more visual attention due to the intensive colour shading. Readers are advised to use caution when trying to interpret health outcomes, including excess mortality in these sparsely-populated counties. Fourth, with reference to data management, we found that most spreadsheet programs are limited in accommodating data in a worksheet (for example, 65,536 rows or records). Conventional spreadsheet programs are insufficient for storing the aggregated data of multiple years required in the present study (N > 195 k). Instead of using spreadsheet programs, we recommend the use of a relational database for data management for spatial-temporal analysis using SatScan.

As observed in Jacquez and Greiling [23,24] of the Journal, several methodological issues involving spatial analysis warrant consideration when using SatScan for cluster analysis. Among other potential limitations are the assumption that the clusters are cylindrically-shaped and the constraints that are attributable to centroids and the edge effects of the scan method. As the authors advised, the scan statistic is but one tool that one may bring to bear on the study of geographic variation in cancer. Nevertheless, th e shape of cluster detection in SatScan may be enhanced over time, and improved methods of utilizing SatScan for cluster analysis are emerging [25,26]. As users of this program, we found that it affords a great opportunity of analyzing the unique geo-demographic composition of Texas data. Particularly appealing is the fact that the program is in the public domain. It provides an opportunity for the integration (as a calculation engine) with other mapping programs, such as EpiMap, freely available from the CDC. Currently, the authors are working to develop an integrated solution using these two programs as a health surveillance system for Texas counties.

Conclusion

Between 1990 and 2001, 4 regions of potential excess breast cancer mortality of statistical significance were detected in Texas counties among non-Hispanic White and Hispanic groups, for the time period between 1990 and 1998. Among all racial groups, both non-Hispanic

Whites in Southeast Gulf Coast Texas and Hispanics in West Texas had the highest mortality burden of this disease, as evidenced by spatial concentration and temporal persistence. There was no evidence that the excess mortality may have persisted through the year 2000 or later in Texas. The excessive occurrence of breast cancer in 3 counties in the Gulf Coast region (for non-Hispanic Whites and Blacks) and in West Texas (for Hispanics) warrants further investigation.

Authors' Contributions

All authors collaborated intensively in this study. CEH contributed to the conceptualization of this study, including visual basic programming and data analysis. HJ and FSM contributed to data interpretation, health disparity analysis and the realignment of the manuscript. All read and approved the final manuscript.

References

1. United States Cancer Statistics: **2000 Incidence Report. National Center for Chronic Disease Prevention and Health Promotion, Center for Disease Control and Prevention.** . November 13, 2003
2. Devesa SG, Blot DJ, Pennello WJ, Hoover G, RN Fraumeni JF Jr: **Atlas of cancer mortality in the United States, 1950–94. Washington, DC: US Govt Print Off; 1999 [NIH Publ No. (NIH) 99-4564].** .
3. Cooper SP, Sigurdson A, Labarthe D, Whitehead L, Downs T, Burau K, Vernon SW, Spitz M, New B: **Assessing the burden of cancer in Texas using vital statistics data.** *South Med J* 1998, **91**:173-81.
4. Zhan FB, Lin H: **Geographic patterns of cancer mortality clusters in Texas, 1990 to 1997.** *Texas Medicine* 2003, **99**:58-64.
5. Zhan FB: **Are deaths from liver cancer, kidney cancer, and leukemia clustered in San Antonio?** *Texas Medicine* 2002, **98**:51-6.
6. **Texas vitalweb ICD-10 Mortality.** . November 29, 2003
7. Newman LA, Mason J, Cote D, Vin Y, Carolin K, Bouwman D, Colditz GA: **African-American ethnicity, socioeconomic status, and breast cancer survival: a meta-analysis of 14 studies involving over 10,000 African-American and 40,000 White American patients with carcinoma of the breast.** *Cancer* 2002, **94**:2844-54.
8. Shavers VL, Harlan LC, Stevens JL: **Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35.** *Cancer* 2003, **97**:134-47.
9. Kulldorff M, and Information Management Services, Inc: **SaTScan™ User Guide. SaTScan v. 4.0: Software for the spatial and space-time scan statistics.** 2003.
10. Census: **Census 2000 Population. US Census Bureau.** [http://factfinder.census.gov]. November 22, 2003
11. **Population Data** *Texas Department of Health* [http://soupfin.tdh.state.tx.us/people.htm]. November 22, 2003
12. US Gazetteers: [http://www.census.gov/tiger/tms/gazetteer/counties.txt]. November 22, 2003
13. Kulldorff M, Nagarwalla N: **Spatial disease clusters: Detection and Inference.** *Statistics in Medicine* 1995, **14**:799-810.

14. Kulldorff M, Athas WF, Feurer EJ, Miller BA, Key CR: **Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico.** *Am J Public Health* 1998, **88:**1377-80.

15. Kulldorff M, Feuer EJ, Miller BA, Freedman LS: **Breast cancer clusters in the northeast United States: a geographic analysis.** *Am J Epidemiol* 1997, **146:**161-70.

16. Roche LM, Skinner R, Weinstein RB: **Use of a geographic information system to identify and characterize areas with high proportions of distant stage breast cancer.** *J Public Health Manag Pract* 2002, **8:**26-32.

17. Song C, Kulldorff M: **Power evaluation of disease clustering tests.** *Int J Health Geogr* 2003, **2:**9.

18. Kulldorff M, Tango T, Park PJ: **Power comparisons for disease clustering tests.** *Computational Statistics & Data Analysis* 2003, **42:**665.

19. Jemal A, Kulldorff M, Devesa SS, Hayes RB, Fraumeni JF Jr: **A geographic analysis of prostate cancer mortality in the United States, 1970–89.** *International Journal of Cancer* 2002, **101:**168-74.

20. Sasco AJ: **Epidemiology of breast cancer: an environmental disease?** *APMIS* 2001, **109:**321-32.

21. Tabar L, Yen MF, Vitak B, Chen H, Smith R, Duffy S: **Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening.** *Lancet* 2003, **361:**1405-10.

22. Sheehan TJ, Gershman ST, MacDougall LA, Danley RA, Mroszczyk M, Sorensen AM, Kulldorff M: **Geographic Assessment of Breast Cancer Screening by Towns, Zip Codes, and Census Tracts.** *Journal of Public Health Management & Practice* 2000, **6:**48-57.

23. Jacquez GM, Greiling DA: **Local clustering in breast, lung and colorectal cancer in Long Island, New York.** *Int J Health Geogr* 2003, **2:**3.

24. Jacquez GM, Greiling DA: **Geographic boundaries in breast, lung and colorectal cancers in relation to exposure to air toxics in Long Island, New York.** *Int J Health Geogr* 2003, **2:**4.

25. Gregorio DI, Samociuk H: **Breast cancer surveillance using gridded population units, Connecticut, 1992 to 1995.** *Ann Epidemiol* 2003, **13:**42-9.

26. Gregorio DI, Kulldorff M, Sheehan TJ, Samociuk H: **Geographic distribution of prostate cancer incidence in the era of PSA testing, Connecticut, 1984 to 1998.** *Urology* 2004, **63:**78-82.