ISSN: 2753-7919
British Journal of Healthcare and Medical Research; Vol. 13 No. 02 (2026) (232-242)
https://doi.org/10.14738/bjhr.1302.20170

Spatial Autocorrelation of Pancreatic Cancer Incidence Across ZIP Codes in Hillsborough County, Florida

Arshita Singh[1], Aarya Satardekar[2], Anusha Parajuli[2], Spuritha Bhandaru[2], Namit Choudhari[3], Rishil Shah[4], Salah Komrojki[5], Benjamin G. Jacob[2]

1. Department of Chemistry, College of Arts and Sciences, University of South Florida, United States of America
2. Samuel P. Bell III College of Public Health, University of South Florida, Tampa, FL, USA
3. School of Geosciences, University of South Florida, Tampa, FL, USA
4. Bellini College of Artificial Intelligence, Cybersecurity and Computing, University of South Florida, Tampa, FL, USA
5. Tampa General Hospital; Moffitt Cancer Institute, Tampa, FL, USA

Abstract: Pancreatic cancer is one of the most lethal malignancies in the United States, with survival outcomes strongly influenced by early detection and underlying social determinants of health. This study examines the spatial distribution of pancreatic cancer incidence across ZIP codes in Hillsborough County and evaluates the relationship between geographic clustering and sociodemographic characteristics. Using spatial analytical techniques, we tested for geographical autocorrelation and identified statistically significant hotspots and cold spots of pancreatic cancer incidence. Global spatial autocorrelation was assessed using Moran's I to determine whether incidence rates were spatially clustered across ZIP codes. Local cluster detection was then performed with Getis–Ord Gi* statistic to identify areas with significantly higher or lower incidence relative to neighboring locations. To explore potential drivers of these spatial patterns, regression analyses were conducted using the ZIP code–level sociodemographic variables. Independent variables included racial composition, poverty rate, educational attainment, and indicators of socioeconomic disadvantages. These variables were evaluated for their association with pancreatic cancer incidence rates to determine whether disparities in disease burden corresponded with specific demographic and socioeconomic characteristics. Regression findings indicated that several sociodemographic variables, particularly income and racial composition, were significantly correlated with increased incidence rates. These findings highlight the importance of integrating spatial epidemiology with sociodemographic analysis to identify communities experiencing disproportionate disease burden. Targeted public health interventions, improved screening awareness, and resource allocation in identified hot spot areas may help address disparities in pancreatic cancer outcomes within Hillsborough County.

Keywords: Pancreatic Cancer, ZIP Codes, Getis–Ord Gi* statistic, Moran's I ,Hillsborough County, Florida.

INTRODUCTION

Pancreatic cancer is among the most commonly diagnosed cancers in the United States, yet it presents with a 5-year survival rate of only 13.3% [1]. In 2025 alone, this roughly translated to 67,440 new diagnoses and 51,980 deaths in the United States [1]. The poor prognosis can be attributed to the later stages of diagnosis and limited treatment options.

Pancreatic cancer is the abnormal growth of cells in the pancreas, an organ that is situated posterior to the lower stomach and synthesizes enzymes for digestion and secretes hormones [2]. The most frequently diagnosed form of pancreatic cancer is pancreatic ductal adenocarcinoma (PDAC), arising in ducts that carry digestive enzymes from the pancreas [2].PDAC is an aggressive form of cancer and can be very resistant to treatment. Some of the risk factors that contribute to the diagnosis of pancreatic cancer include chronic pancreatitis, diabetes mellitus, tobacco use, obesity, and genetic predisposition [3]. Despite these risk factors, the symptoms are nonspecific and can often be misinterpreted for less severe conditions and go undetected, contributing to disease progression [2].

Early detection is vital for improving survival outcomes. If pancreatic cancer is detected early, localized cancers allow for a 44% five-year relative survival rate [4]. However, if discovered later, regional cancers have a five-year relative survival rate of 16%, and distant cancers present a survival rate of 3% [4].

Current efforts in Florida to improve early detection include risk assessment tools and advanced diagnostic imaging techniques at Moffitt, Baycare, and UF, amongst others. However, accessibility barriers, including a lack of insurance coverage, health disparities, and a failure to prioritize high-risk populations, remain vital to early diagnosis and treatment.

Health disparities play a significant role in the incidence and outcomes of pancreatic cancer. In the United States, the Black or African American population experiences significantly higher mortality and prevalence rates compared to other ethnic groups [4]. Furthermore, Fuko et al. found that Black/African American and Asian/Pacific Islander populations exhibited higher odds of mortality compared to other racial groups for pancreatic cancer [5]. These disparities indicate that socioeconomic, environmental, and access factors may play a role in the disease burden and mortality.

Regression and geographic analysis are important resources for understanding the distribution of cancer and identifying high-risk populations. While other papers have utilized regression analysis and geographical processing, they lack specificity to pancreatic cancer or geographical analysis of specific counties. For instance, Vieira and others examined the association between location and the prevalence of bladder, kidney, and pancreatic cancer on upper Cape Cod using GIS and spatial-cluster analysis [6]. In the same way, Amini and others examined the global long-term trends in pancreatic cancer incidence and mortality based on location and spatial-cluster analysis. Nonetheless, the lack of specificity in these studies impedes targeted public health interventions [7]. To date, no studies have utilized spatial and regression analysis on pancreatic cancer incidence in Hillsborough County, Florida.

This study aims to address this gap by using spatial analysis to test for geographical autocorrelation and hot and cold spots for pancreatic cancer incidence. Moreover, we utilized regression analysis tests to identify sociodemographic variables that correlate with increased incidence in Hillsborough County for each zip code. By identifying the sociodemographic variables and spatial patterns, the research seeks to promote targeted intervention strategies to improve early detection and reduce disparities among individuals in Hillsborough County.

METHODOLOGY

Study Area

Within Florida, Hillsborough County houses approximately 1.5 million residents, making it the 2nd largest county by population and the 13th-largest county by land coverage [8]. There are approximately 244 new cases of pancreatic cancer every year in Hillsborough County, with 72% of the cases being in the last stage, highlighting the need for interventions related to pancreatic cancer [9]. Hillsborough County has a mixture of urban, farmland, and rural land cover according to the enriched Volume 6-Issue 5 mix that allows for the diverse variability that can be utilized for statistical modeling for pancreatic cancer.

Study Site Map

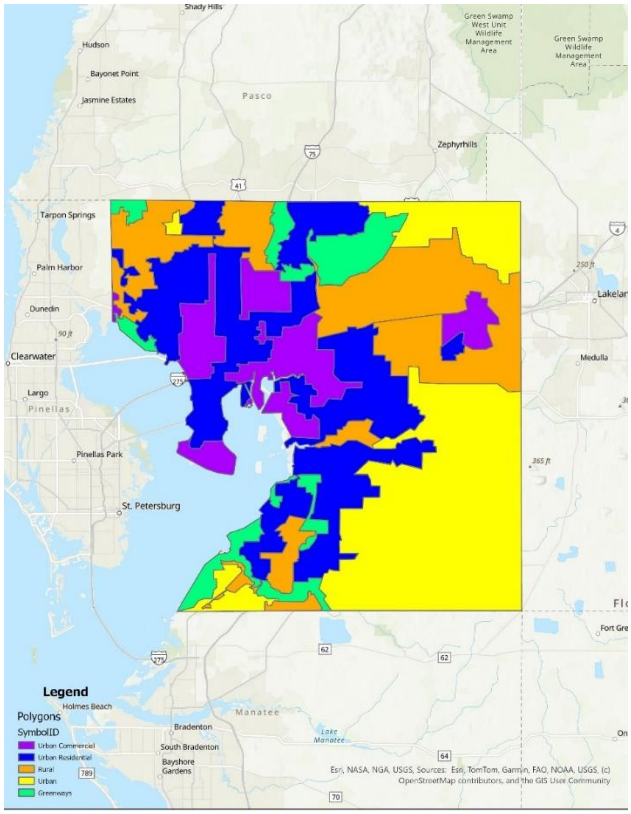

Figure 1: A Map of Hillsborough County, Florida, by Land Use Land Cover (LULC)

Population Stratification

The unit of analysis of the PC covariates was the Zip Code Tabulation Area (ZCTA), which approximates U.S. Postal Service zip codes and is constructed by the U.S. Census Bureau. ZCTA boundary shapefiles. All analyses were conducted at the ZCTA level to ensure compatibility between spatial boundaries and publicly available census-derived covariates.

The study employed a count variable Poisson model to quantify land use land cover (LULC) and socio-demographic co-variants associated with pancreatic cancer. Literature and census data were utilized to create a population stratification for the entire county of Hillsborough, which included all 55 zip codes (Table 1). We created the conversion of

(Incidence: Population), using the ratio ≈ 244/1.6 million. Subsequently, we quantified potential cases at the zip code level employing the following equation $\frac{(244)(zipcode\ population)}{1.6\ million}$.

The known incidence of cases of pancreatic cancer was employed along with covariants and population stratification in a Poisson regression model framework to generate a parameter estimator hierarchy. Finally, a second-order eigenfunction decomposition was utilized to cartographically outline potential hot and cold spots with a local Moran's index. The statistical analysis through the statistical modeling and geographical analysis is essential for clinical and epidemiological significance.

Table 1: Table of all the ZCTA values and population stratification data

Zip code	Population	Vulnerability Index
33647	74919	11.38573152
33511	58987	8.964483578
33578	55724	8.46859279
33612	50697	7.704620068
33617	48987	7.444744724
33615	48849	7.423772328
33614	46844	7.119064688
33610	45622	6.9333526
33579	43456	6.604177164
33624	41661	6.33138404
33604	40328	6.128802851
33619	39272	5.968318428
33613	37354	5.676832515
33594	35824	5.444312471
33611	34877	5.300393201
33570	32620	4.957388142
33626	30736	4.671069342
33547	30431	4.624717307
33556	30386	4.617878482
33510	30103	4.574869872
33558	29098	4.422136118
33625	28818	4.37958343
33596	28558	4.340070219
33573	28491	4.329887969
33618	28416	4.318489928
33598	27978	4.251925366
33584	27863	4.234448369
33563	26874	4.084146196
33629	26625	4.046304699
33569	26189	3.980044085
33607	24257	3.686430538
33572	23846	3.623969271
33634	22927	3.484305271
33566	22842	3.47138749
33606	22546	3.42640322
33635	19917	3.026863876

33603	19545	2.97032959
33559	19226	2.921849921
33609	18411	2.797991204
33549	18016	2.737961519
33565	17767	2.700120022
33602	17631	2.679451574
33637	17507	2.660606812
33534	17479	2.656351543
33527	16342	2.483557235
33605	16166	2.456809831
33616	14918	2.267146422
33567	12499	1.899521594
33592	11295	1.716545036
33548	7639	1.160928511
33620	6451	0.980383535
33621	2309	0.350907701
33503	401	0.060941528
33530	18	0.00273553

Data Sources and Covariate Construction

The dependent variable in the PC estimator determinant dataset consisted of count data aggregated to the ZCTA level (e.g., incident cases, events, or service utilization counts). A population offset term was constructed using total ZCTA population estimates to account for exposure differences across areas. All tabular census data were joined to ZCTA shapefiles using unique ZCTA identifiers in a GIS environment (e.g., ArcGIS). Spatial processing included: 1. Projection to a common coordinate reference system. Verification of topology 3. Removal of ZCTAs with incomplete covariate data. Assessment of multicollinearity via Variance Inflation Factors (VIF).

Poisson Regression Modeling

Because the outcome variable consisted of count data with non-negative integer values, a generalized linear model (GLM) with a Poisson distribution and log link function was employed. The model was specified as:$\log(\mu_i) = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ik} + \log(Pop_i)$where:

- μ_i= expected count in ZCTA i
- X_{ik}= PC covariate $k$for ZCTA i
- β_k= regression coefficient
- Pop_i= population offset

Model adequacy was evaluated through 1. Deviance and Pearson goodness-of-fit statistics 2. Overdispersion testing (ratio of deviance to degrees of freedom)3. Akaike Information Criterion (AIC) for model comparison.

To determine parameter hierarchy (relative importance of covariates), we used multiple approaches:

1. **Standardized Coefficients**: Because all predictors were standardized, the magnitude comparison of ß coefficients directly reflected relative influence.
2. **Likelihood Ratio Tests**: Nested models were compared by sequentially removing conceptual blocks (socioeconomic → sociodemographic → racial) to evaluate domain-level contributions.
3. **Partial Deviance Explained**: The proportional reduction in deviance attributed to each variable or domain was calculated.
4. **Dominance Analysis (Optional Advanced Approach)**: Relative importance metrics were computed to assess each predictor's average contribution across model subsets.

Hierarchical interpretation focused on both statistical significance ($p < 0.05$) and effect magnitude.

Spatial Autocorrelation and Hot/Cold Spot Analysis

To evaluate whether the distribution of the outcome and PC covariates exhibited spatial clustering beyond random expectation, spatial autocorrelation analyses were conducted on the zip code-stratified covariates. A spatial weights matrix (W) was constructed using Queen contiguity (shared edges or vertices). Weights were row-standardized to ensure comparability across ZCTAs. Global Moran's I was computed to assess overall clustering of: The outcome variable and model residuals from the Poisson regression. Statistical significance was assessed using 999 Monte Carlo permutations. Significant positive Moran's I indicated clustering; negative values indicated dispersion.

Local Indicators of Spatial Association (LISA) were calculated to identify spatial clusters at the ZCTA level. Additionally, the Getis–Ord Gi* statistic was computed to produce statistically significant hot spot (high-value clustering) and cold spot (low-value clustering) maps. For each ZCTA: $G_i^* = \frac{\sum_j w_{ij}x_j - \bar{X}\sum_j w_{ij}}{S\sqrt{\left[\frac{n\sum_j w_{ij}^2 - (\sum_j w_{ij})^2}{n-1}\right]}}$

where:

- w_{ij}= spatial weight between ZCTAs
- x_j= value of interest
- n= total number of ZCTAs

Z-scores and corresponding p-values were used to classify:

- High-high clusters (hot spots)
- Low-low clusters (cold spots)
- Spatial outliers (high-low, low-high)

Spatial clustering of the raw outcome counts, regression residuals, and predicted values was examined to determine whether the Poisson model adequately accounted for

spatial dependence. Persistent clustering in residuals indicated potential spatial processes not captured by covariates and suggested the need for spatial regression extensions (e.g., spatial lag or conditional autoregressive models).

All spatial analyses were conducted using GIS software (e.g., ArcGIS Pro, QGIS) and statistical modeling was performed in R (packages: *spdep*, *sf*, *MASS*, *car*) or Python (e.g., PySAL).

RESULTS

To analyze the socio-demographic covariants associated with pancreatic cancer, a Poisson regression analysis was utilized, but it reported that none of the covariates were statistically significant (p-values > 0.05) due to the variance inflation factors being greater than 10, illustrating many multicollinearity issues. Then, a negative binomial was employed, but none of the covariates were statistically significant due to the same severe multicollinearity problems with high VIF values, as many of the variables are highly correlated. A stepwise backward elimination with Akaike Information Criterion (AIC) was utilized to remove variables that did not allow for a significant model fit. The final model determined Asian population and poverty population as the two covariates with the lowest AICs, identifying them as the most important predictors of pancreatic cancer in Hillsborough County (Table 2). Finally, the parsimonious model analyzed the sociodemographic variables determined by the stepwise backward elimination to determine the statistical properties and interpretability of the covariates. Population poverty was highly statistically significant with a $\beta = 3.844\times10^{-5}$ and $p < 0.001$ (Table 3). Therefore, areas with higher rates of poverty had significantly higher incidences of pancreatic cancer. For the Asian population, there was a marginal negative association with $\beta = -6.059\times10^{-5}$, $p = 0.0625$, but it is not statistically significant (Table 3). With this model, the multicollinearity was effectively reduced with the VIF values of 1.71 for both variables (Table 4).

Table 2: Results of the Stepwise Backward Elimination

Variable		Df	AIC
<none>			186.09
Asian		1	189.85
Poverty Population		1	210.28

Table 3: Results of the parsimonious model to determine statistical significance

Variable	Estimate	Std. Error	z-value	Pr(>|z|)
Intercept	3.111e-01	1.711e-01	1.818	0.0690
Asian	-6.059e-05	3.253e-05	-1.863	0.0625
Poverty Population	3.844e-05	5.303e-06	7.249	4.21e-13

Table 4: VIF values with the final parsimonious model

Variable	VIF value
Asian	1.711266
Poverty Population	1.711266

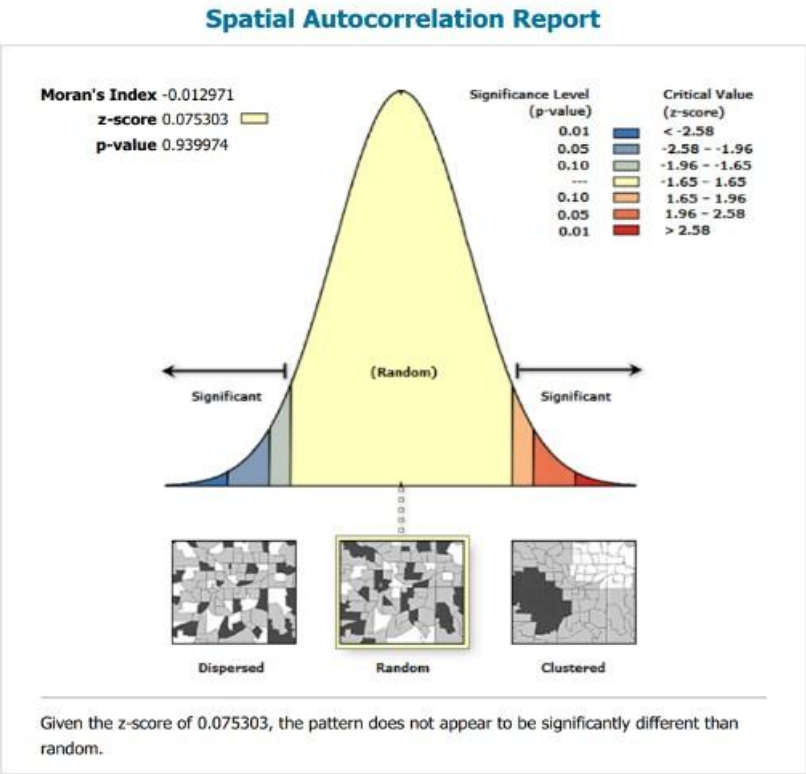

Figure 2: Spatial Autocorrelation Report

The spatial autocorrelation determines the z-score of 0.075303, indicating that the pattern of geospatial correlation with pancreatic cancer is not significantly different than the random spatial location.

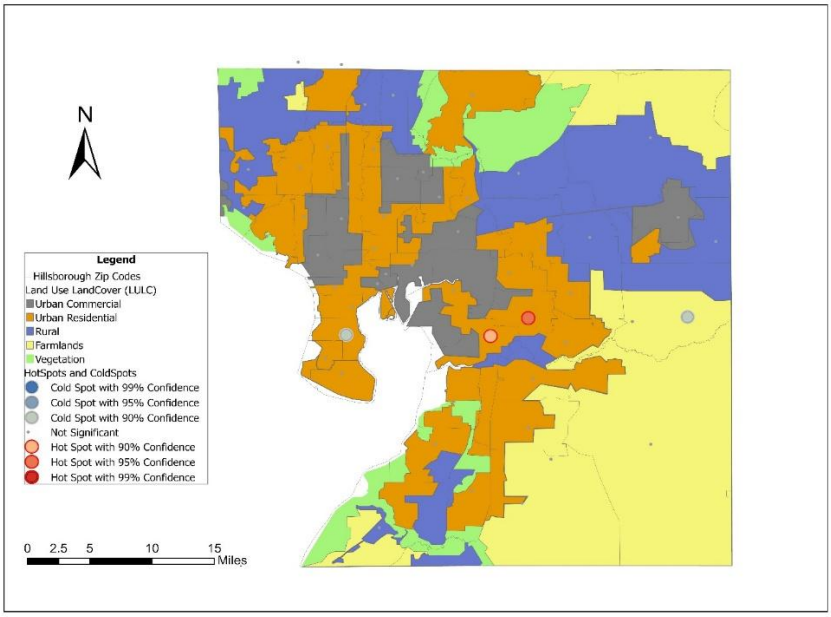

Figure 3: Hot and cold maps with geographical variation

The map illustrates different zip codes marked based on geographical classifications, such as urban commercial, urban residential, rural, farmlands, and vegetation. The map

also illustrates hot and cold spots that have been coded for size to indicate the confidence levels and color to indicate if they represent more cases of pancreatic cancer or fewer (red or blue, respectively).

DISCUSSION

The statistical modeling framework, which included Poisson regression, stepwise backward selection, and negative binomial regression, identified poverty as a statistically significant covariate. This finding indicates that geographic areas with a higher proportion of residents living in poverty exhibited an increased incidence of pancreatic cancer cases (Table 3). The model also identified the Asian population as having a negative association with pancreatic cancer incidence; however, this relationship did not reach statistical significance (Table 3). Spatial analysis indicated that pancreatic cancer cases were not spatially autocorrelated, suggesting a pattern consistent with random spatial dispersion of pancreatic cancer incidence across the study area (Figure 2). The hot spot analysis further identified 33578 and 33511as a statistically significant hot spot location with 90% and 95% confidence, respectively (Figure 3).The zip codes 33567 and 33611 were identified as cold spots with 90% confidence (Figure 3).

Within the existing literature on pancreatic cancer epidemiology, there is evidence suggesting that poverty may influence cancer risk through mechanisms related to education and health behavior. It is well established that poverty limits educational attainment, whereas higher levels of education can provide opportunities that reduce exposure to socioeconomic disadvantage [10]. Education also plays a critical role in shaping health behaviors that directly influence cancer risk. For example, smoking is a well-established risk factor for pancreatic cancer, yet individuals with lower levels of educational attainment are more likely to smoke and experience poorer outcomes when attempting smoking cessation. Similarly, other modifiable behavioral factors—including alcohol consumption, poor dietary patterns, and obesity—have been associated with increased cancer risk [11]. In this context, education may act as an important mediating factor that reduces engagement in unhealthy behaviors, thereby suggesting a plausible pathway linking poverty to increased cancer risk [12].

The hot spot analysis of pancreatic cancer incidence also illustrates clusters of elevated incidence within two urban areas. This spatial pattern of increased pancreatic cancer rates in urban settings has been reported in global epidemiological research, where higher levels of urbanization and affluence have been associated with increased alcohol consumption and higher prevalence of chronic pancreatitis, both of which are recognized risk factors for pancreatic cancer [13].

Pancreatic cancer is widely recognized as a disease associated with aging, with incidence rates increasing substantially among older populations[14]. Additionally, prior epidemiological studies have reported that Black/African American individuals experience higher incidence rates of pancreatic cancer compared to other racial and ethnic groups. However, the statistical analysis conducted for Hillsborough County did not identify this racial group as being significantly associated with increased pancreatic cancer incidence within the present model [5].

This initiative proposes the development of a geospatial artificial intelligence (AI) platform integrated with a smartphone mobile application can deliver targeted health messaging to vulnerable populations residing in statistically identified pancreatic cancer hot spots. The platform can integrate cancer incidence data with sociodemographic indicators—including race, income, educational attainment, and healthcare access—to identify vulnerable communities with elevated risk. A geospatial AI engine in the platform can then generate location-specific risk profiles and prioritizes ZIP codes where targeted public health outreach may have the greatest impact. Through the accompanying smartphone mobile application, residents in identified hot spot areas can receive tailored social messaging focused on pancreatic cancer awareness, risk factor education, symptom recognition, and guidance for accessing local screening and healthcare services. The system also supports push notifications, interactive educational content, and geolocated resources such as nearby clinics and support services. By combining spatial epidemiology, AI-driven analytics, and mobile health communication, this approach enables precision public health interventions at the neighborhood level. Targeted digital outreach can improve awareness, promote earlier clinical engagement, and reduce disparities in pancreatic cancer outcomes among high-risk populations in Hillsborough County.

One limitation of this study is the potential for ecological fallacy, as pancreatic cancer incidence was analyzed at the population level rather than at the individual level. As a result, associations observed at the aggregated population scale may not necessarily reflect relationships present at the individual level within each ZIP code. Future research could extend this analysis to the statewide level, enabling a more comprehensive assessment using a larger sample size and a broader set of predictive covariates associated with pancreatic cancer incidence. Another important direction for future research involves the development of targeted clinical interventions aimed at populations experiencing poverty, which may facilitate earlier detection of pancreatic cancer.

CONCLUSION

Given the high mortality and increasing incidence associated with pancreatic cancer, this study identifies populations residing in areas characterized by poverty and urban concentration as potential targets for focused public health interventions..

Using spatial epidemiologic methods, including geographic autocorrelation and local cluster detection such as Moran’s I and Getis–Ord Gi* statistic, ZIP code–level pancreatic cancer incidence data can be analyzed to identify significant hot spots and cold spots across Hillsborough County. By applying advanced statistical modeling techniques and geospatial analysis, this research highlights geographic areas that may benefit from enhanced surveillance and preventive strategies.

Future clinical and public health research focusing on these high-risk populations may support the development of preventive screening programs, targeted interventions, and early detection strategies, which could ultimately contribute to reducing pancreatic cancer mortality.

REFERENCES

[1] "Cancer of the Pancreas - Cancer Stat Facts," SEER. Accessed: Feb. 17, 2026. [Online]. Available: https://seer.cancer.gov/statfacts/html/pancreas.html

[2] "Pancreatic cancer - Symptoms and causes," Mayo Clinic. Accessed: Feb. 17, 2026. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/pancreatic-cancer/symptoms-causes/syc-20355421

[3] A. Mannucci and A. Goel, "Advances in pancreatic cancer early diagnosis, prevention, and treatment: The past, the present, and the future," *CA. Cancer J. Clin.*, vol. 76, no. 1, p. e70035, 2026, doi: 10.3322/caac.70035.

[4] "Pancreatic Cancer - UF Health." Accessed: Feb. 17, 2026. [Online]. Available: https://ufhealth.org/conditions-and-treatments/pancreatic-cancer

[5] C. D. Fuko, H. M. Magacha, G. Noah, and O. V. Ikwuka, "Ethnic/Racial Disparities in Pancreatic Cancer Mortality Across the United States: A National Inpatient Sample Database Analysis," *Cureus*, vol. 17, no. 1, p. e77389, doi: 10.7759/cureus.77389.

[6] V. Vieira, T. Webster, J. Weinberg, and A. Aschengrau, "Spatial analysis of bladder, kidney, and pancreatic cancer on upper Cape Cod: an application of generalized additive models to case-control data," *Environ. Health*, vol. 8, no. 1, p. 3, Feb. 2009, doi: 10.1186/1476-069X-8-3.

[7] M. Amini, M. Azizmohammad Looha, S. Rahimi Pordanjani, H. Asadzadeh Aghdaei, and M. A. Pourhoseingholi, "Global long-term trends and spatial cluster analysis of pancreatic cancer incidence and mortality over a 30-year period using the global burden of disease study 2019 data," *PLOS ONE*, vol. 18, no. 7, p. e0288755, Jul. 2023, doi: 10.1371/journal.pone.0288755.

[8] "Census Bureau Data." Accessed: Mar. 12, 2026. [Online]. Available: https://data.census.gov/

[9] "State Cancer Profiles > Incidence Rates Table." Accessed: Mar. 12, 2026. [Online]. Available: https://www.statecancerprofiles.cancer.gov/incidencerates/index.php?statefips=12&areatype=county&age=001&cancer=040&stage=211&race=00&sex=0&ruralurban=0&year=0&sortVariableName=cirank&sortOrder=asc&type=incd&output=0

[10] H. Ferguson, S. Bovaird, and M. Mueller, "The impact of poverty on educational outcomes for children," *Paediatr. Child Health*, vol. 12, no. 8, pp. 701–706, Oct. 2007, doi: 10.1093/pch/12.8.701.

[11] R. Ranganath and Q. Chu, "Global trends in pancreas cancer among Asia-Pacific population," *J. Gastrointest. Oncol.*, vol. 12, no. Suppl 2, pp. S374–S386, Jul. 2021, doi: 10.21037/jgo-20-118.

[12] S. N. Patel *et al.*, "The Impact of Social Determinants on Pancreatic Cancer Care in the United States," *Cancers*, vol. 17, no. 12, p. 1898, Jun. 2025, doi: 10.3390/cancers17121898.

[13] J. Zottl, C. G. Sebesta, E. Tomosel, M.-C. Sebesta, and C. Sebesta, "Unraveling the Burden of Pancreatic Cancer in the 21st Century: Trends in Incidence, Mortality, Survival, and Key Contributing Factors," *Cancers*, vol. 17, no. 10, p. 1607, May 2025, doi: 10.3390/cancers17101607.

[14] "Disparities in cancer outcomes across age, sex, and race/ethnicity among patients with pancreatic cancer - Nipp - 2018 - Cancer Medicine - Wiley Online Library." Accessed: Mar. 18, 2026. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1002/cam4.1277