INTRODUCTION

Cervical cancer was a leading cause of cancer death among American women before screening began in the 1940s [1]. Approximately 12,900 American women will be diagnosed with cervical cancer during 2015, and 4,100 will succumb to this disease during the same year [2]. During 2000, California showed the second highest incidence rate for cervical cancer in the US [3], with 2,663 deaths from cervical cancers among residents in California during 2006-2011 [6]. Overwhelmingly, cervical cancers arise from premalignant dysplasia, which can progress to cancer. Since the 1950s [6], the Papanicolaou (Pap) test has been used for detection of premalignant cervical dysplasia and cancer [3,6,7] and to improve survival through early treatment [8,9]. The majority of cervical cancers in the US are epidermoid carcinomas (67%), with the remainder classified as a deno carcinomas (27%) or other histologies that include specified and unspecified carcinomas or sarcomas [10]. Approximately 99 percent of cervical cancers result, in part from one of 13 human papilloma virus (HPV) subtypes [11], with two vaccines available for the most common subtypes [12]. In spite of the effectiveness of HPV vaccination, the value of Pap testing for detection of premalignant cervical dysplasia and early cervical cancer will extend into the foreseeable future [13-15].

Race and ethnicity are frequently used as predictors of late-stage cervical cancer [16,17], while researchers have suggested that this effect results substantially from low SES [16,18,19]. Low SES, which is theoretically modifiable, predicts inadequate Pap testing [20-22], higher incidence rates [20,23-26] and later stage at diagnosis [8,10,19] for cervical cancer, with these outcomes predicted by large area geographic units. Intensified targeting of screening to high risk regions of California may improve cervical cancer control and prevention.

Various researchers have used community SES indices ranging from single variable measures to those using multiple dimensions including income [20], education [51], poverty [21,51,52] and other social or economic characteristics. Three indices measuring ecologic SES quintiles at the census blockgroup...
(BG) level have been developed for use in the diverse California population. Among these, the prototype multidimensional BG level SES quintile index was based on a principal component regression model using seven, Year 1990 Census long form variables that included education, median income, percentage living below the poverty level, median rent, median house value, proportion with a blue-collar job and proportion in the workforce without a job that were older than age 16 years [27]. Similar methods used block group data from the Year 2000 Census, updating the original index for 1996-2005. The most current SES index, applicable to years 2006-2011, was derived using the same component variables and methods as the earlier indices, used Year 2010 Census BGs and data from the American Community Survey [28].

The California Cancer Registry (CCR), operated continuously since 1988, consists of the three most populated Surveillance Epidemiology and End Results (SEER) program registries and includes data for malignant neoplasms diagnosed among more than 37 million California residents [10]. Variables available in the CCR include age, sex, community SES quintiles, demographic information, cancer anatomic site and histology, behavior and Year 2010 Census BG of residence at diagnosis.

**OBJECTIVES**

This study sought to use epidemiologic methods and geographic information systems (GIS) tools to determine whether sub-county aggregations of cervical cancers existed in California and to map geographic areas depicting unique cervical cancer occurrence. Additionally, this research sought to map sub-county areas of California exhibiting unique SES patterns and to map and test spatial cross-correlations between aggregations of unique cervical cancer occurrence and SES, seeking to enhance control and prevention strategies.

**MATERIAL AND METHODS**

This population-based, cross-sectional study identified geographic aggregations of cervical cancer occurrence measured at the Year 2010 Census BG level in California for 2006-2011 using Statistical Analysis Software (SAS) version 9.4 [29] and Arc GIS version 10.3 [30]. Variables used in this analysis included age at diagnosis, year of diagnosis (2006-2011), histology for invasive cervical epidermoid carcinoma (M-8050-8052, 8070-8076, 8082-8084) and adeno carcinomas (M-8140, 8144-8145, 8147, 8200, 8210, 8240, 8245, 8246, 8255, 8260, 8263, 8310, 8323, 8380, 8384, 8430, 8441, 8460-8461, 8480, 8482, 8490, 8501, 8523, 8542, 8560, 8570, 8574) [31], an ecologic SES quintile index (1-lowest, 5-highest) [28] for each California, Year 2010 Census BG, regardless of other demographic characteristics. Age-, race/ethnicity- and marital status-cruide incident ratios (IR) for cervical cancer were computed as the ratio of observed cases, divided by the expected count for each of the 23,212 Year 2010 Census BGs in California. The expected number for each BG was computed using indirect standardization [32] by applying the statewide average risk to the number of women residents, age 20 years or older in each California Year 2010 Census BG.

BG-level IRs was mapped using a spatial smoothing method [42,43] to improve statistical reliability for sparsely populated, sub-county areas. IR values were estimated for individual points that were spaced every two kilometers forming a grid that encompassed the entire state. This adaptive kernel, sub-county estimation method aggregated observed and expected counts of cervical cancers into larger, homogeneous geographic units that included a minimum of 5,000 females older than age 20 years. This spatial smoothing method ensured adequate sample size for mapped homogeneous geographic units. GIS functions were used to create six ordinal categories for cervical cancer IRs, which were then mapped [33,34].

We utilized a three-step process to formally evaluate spatial clustering in the geographic distribution of cervical cancer [30]. The first step involved calculating cervical cancer IRs for each collection of eight nearest neighbor BGs using a spatial weights matrix [35]. In step two, the global Moran’s I coefficient was used to assess global spatial auto correlation between each BG, with the closest eight BG neighbors [36]. Step three involved calculating local indicators of spatial association (LISA) [37] to identify clusters of BGs exhibiting IR values that differed significantly from statewide average values. False Discovery Rate (FDR) [38] procedures were used to correct for multiple testing and spatial dependency [39] when reporting statistical significance findings.

A choropleth map of California was constructed depicting the geographic distribution of SES quintile levels measured for Year 2010 Census BGs ((Figure 2) Map A). The cross-LISA index measuring cross-correlation between two standardized spatial variables [40] was computed and plotted using Geo DA software (version 1.0.1) [41]. This index identified geographic areas depicting unique local associations between continuous IR and SES matrix levels. County borders were shown on each map to provide geographic reference.

**RESULTS**

During 2006-2011, 8,413 California residents were diagnosed with cervical cancer. Among these, 337 (4.0%) consisted of non-specific cervical cancer types, 18(0.2%) were some form of basal cell carcinoma and 106 (1.3%) sarcomas and were excluded from study. In addition to these data for non-epidermoid carcinomas and non-adeno carcinomas, subjects missing information for age (n=179 [2.1%]), cases less than age 20 years at diagnosis (n=301 [3.6%]), cases that could not be geo coded to the BG level (n=57 [0.7%]) and 183 (2.2%) cases located in BGs having fewer than five counts were excluded from analysis. Data for the remaining 7,232 eligible study subjects were geo coded to the 5,933 California census block groups having at least one cervical cancer case diagnosed during the six-year study period.

Table (1) presents the distribution of cervical cancers and BG counts according to each of six ordinal IR categories used to classify spatially smoothed data. Among the 7,232 California women diagnosed with cervical cancer, 1,642 (23%) occurred in areas having IR values ranging from greater than 1.25 to 4.77 (the three highest IR categories) at diagnosis. Data in Table 1 showed that 1,260 (21.2%) of the 5,933 California Year 2000 Census BGs having at least one cervical cancer case, were classified in the three highest IR categories. Table 2 presents counts and percentages of total cervical cancer cases and BG categories, according to SES quintiles. These findings indicate that 48
percent of cervical cancer cases occurred in the two lowest SES categories, representing 47 percent of the California census BGs with cervical cancers.

Figure (1) presents a reference map of California depicting county names and boundaries (Figure 1 Map A), a spatially smoothed, hierarchal map of cervical cancer incident ratio categories (Figure 1 Map B) and a map of LISAIR clusters assessed according to statistical criterion ($p < 0.01$) (Figure 1 Map C) [44]. This descriptive map of spatially smoothed ratios reflects differences in sub-areas of the 58 California counties (Figure 1 Map B) with the highest cervical cancer IR levels occurring in eastern San Diego, central Imperial, south central San Bernardino, west central Los Angeles, western Ventura, central Kern, central Kings, south Tulare, central Fresno, western Merced and north western Madera counties. Northern California counties that showed highest IR values include Humboldt and Siskiyou Counties (Figure 1. Map B). Moran’s I [36], a global spatial autocorrelation test, identified a tendency for clustering of cervical cancer IRs within adjacent census BGs (Moran’s $I = 0.738$; $p < 0.001$). Results for the local version of Moran’s I, LISA findings [44] revealed clusters of cervical cancer in areas of eastern Imperial and San Diego counties, desert areas of western San Bernardino County, central Kern and sporadic areas of Los Angeles counties. Additional areas of higher than average IR clusters were evident in Tulare, Kings, Fresno, south Monterey, Madera, Mariposa, Merced, Tuolumne, Butte, Lake, Mendocino, Trinity, Humboldt and Siskiyou Counties (Figure 1 Map C).

(Figure 2) presents a map depicting SES quintile levels in California ((Figure 2 Map A) and a map depicting significant BG clustering ($p < 0.01$) for cross-correlations between IR and SES measures for each BG cluster [35] computed using a cross-correlation coefficient [40]. The geographic distribution of SES quintiles in California shows that the highest levels (depicted in blue) are predominantly located along the California coastline from Sonoma County southward, extending inland along the Interstate 80 corridor from the San Francisco to Sacramento and to mountain areas of eastern Placer and Eldorado counties. In southern California, high SES areas were also identified in coastal regions of Los Angeles, Orange, and San Diego Counties. Areas depicting the lowest SES quintile levels (red) include broad areas of eastern San Bernardino, Riverside and Imperial counties and wide areas of the California central valley that encompass parts of Kings, Kern and Fresno counties. Far northern California counties that depict large areas classified in the lowest SES quintile include northern Trinity, north central and southeastern Humboldt, north central Modoc and north eastern and northwestern Siskiyou counties ((Figure 2) Map A).

Table 1: Counts (n) and percentages (%) of cervical cancer cases and block groups by incident ratio (IR) categories.

<table>
<thead>
<tr>
<th>SES Categories</th>
<th>Cancer Cases</th>
<th>Census Block Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>1</td>
<td>1,890</td>
<td>26.13%</td>
</tr>
<tr>
<td>2</td>
<td>1,586</td>
<td>21.93%</td>
</tr>
<tr>
<td>3</td>
<td>1,440</td>
<td>19.91%</td>
</tr>
<tr>
<td>4</td>
<td>1,286</td>
<td>17.78%</td>
</tr>
<tr>
<td>5 (highest)</td>
<td>1,030</td>
<td>14.24%</td>
</tr>
<tr>
<td>Total</td>
<td>7,232</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Figure (2B) presents a map of the cross-correlation between statistical IR clusters and SES levels, based on the cross-correlation (LISA) index, depicting the degree to which cervical cancer IRs for a particular BG is related to the SES score for the neighboring eight BGs. Low SES and high IR areas (red) are evident in eastern Imperial, north eastern San Diego, wide areas of San Bernardino Counties, eastern Kern, south-central Los Angeles, Tulare, most of Kings, areas of Fresno, an east to west band through central Monterey, much of Madera, western Merced and along the state highway 99 and Interstate 5 corridors through Stanislaus, San Joaquin and Sacramento counties. Areas of low SES and high IR in the far northern part of California include parts of Glenn, Lake, Tehama, eastern Mendocino, southern Trinity, northern Humboldt and areas of Siskiyou and Modoc counties. There were other locations (displayed in blue) where the SES versus IR correlation also followed the expected direction of high SES and low IR. Figure (2C) presents a cross-correlation scatter plot, representing the cross-correlation between the two standardized spatial variables across BGs ($I = 0.287; p < 0.001$).

DISCUSSION

In 2000, California ranked number two for cervical cancer incidence rates in the U.S. [3]. Every cervical cancer case representing failure of early detection (screening) [7] and treatment of premalignant dysplasia [45] and early-stage disease [7,45], identifying the tragedy of this ranking. In previous analyses, researchers identified roles for age, race/ethnicity, SES and marital status as predictors of late-stage cervical cancer diagnosis in California [16,19,46-49], while findings reported here identify geographic location in California depicting elevated cervical cancer incident ratios and case clustering.

Other investigators found that U.S. counties having low screening rates, tend to have higher cervical cancer incidence rates [50-52], more advanced stage diagnosis [50,51,53] and higher than average mortality rates [6,50]. While these findings reveal unique patterns of cervical cancer incidence, stage at diagnosis and risk of death at county levels, the large size and lack of demographic homogeneity for county geographic units in California, limit the value of these earlier findings for targeted cancer control and prevention programs.

Using census BGs, rather than counties, this investigation utilized Local Moran’s I, a measure of geographic correlation, to collapse neighboring BGs having similar cervical cancer IR levels,
revealing clustering at sub-county levels. While our findings of significant geographic patterns in occurrence of cervical cancer are like those reported at county levels in the U.S. [50-52], use of sub-county geographic units depicting homogeneously higher than average cervical cancer IRs provides opportunity for application of focused control and prevention strategies.

In contrast with previous studies, this current investigation identified unique cervical cancer incidence and socioeconomic status patterns measured at the census BG level; the smallest economically homogeneous census unit having cancer occurrence and SES data available. Incorporation of robust GIS analytic techniques, together with population-based California Cancer Registry data revealed patterns of unique cervical cancer occurrence in geographic units that can be targeted for cancer control.

Previous studies showed patterns for lower cervical cancer screening among residents of low SES counties [54], with a tendency for higher incidence rates measured for counties having low SES [55]. Findings reported here, provide map patterns depicting unique incident ratios, statistical LISA-derived IR clusters, SES quintiles and geographic areas depicting uniquely high or low Moran’s I values for spatial correlations between LISA IR cluster and SES values. This general relationship of moderate, negative cross-correlation Figure (2C) presents a cross-correlation scatter plot and suggests spatial correlations between cervical cancer incidence and lower SES quintiles in California.

Sub-county geographic areas depicting unique high LISA-derived IRs clusters, statistically elevated IR clusters and correlations between IR clusters and SES are identified in these results. Although these findings are consistent with county-level evidence reported by other investigators [22,50,51,55], they extend this evidence by revealing geographic associations between cervical cancer incidence rates and low SES at sub-county levels. Combining contemporary population-based cancer registry data for California with, Census 2010 denominators and advanced GIS statistical methods, these new findings provide an intuitive means of identifying sub-county areas of California.
that may benefit, disproportionately, from targeted intensifies cervical cytology screening in the foreseeable future.

**LIMITATIONS**

Exclusions of cervical cancer cases not classified by cell morphology, missing age at diagnosis, not geo coded to the BG level and subjects in BG clusters having fewer than five cases (n=756) has potential to introduce systematic error in our findings. Nevertheless, data for 90.5% (7,232/7,988) of the eligible targeted study subjects were included in this study, minimizing the magnitude of differential bias resulting from missing data. Use of the multidimensional ecologic SES index relies on average SES scores computed for census BGS of residence that will differ from individual SES measures for some individuals. Nevertheless, the multidimensional character of the CCR SES index includes a principal component derived composite of seven community SES dimensions that arguably provides better portrayal of Pap screening barriers than a single component SES measure in individuals. Details regarding derivation and validation of the CCR SES index are available from the Cancer Prevention Institute of California [28]. Methods used to form the adaptive kernel and LISA cluster population (geographic) units rely on weighted averages of census BG measures when forming BG clusters. While this approach is expected to enhance specificity for identifying uniquely high and low geographic IR patterns, averaging disparate results for adjacent BGS contained within clusters tends to reduce sensitivity for identifying uniquely high or low effects. This potential loss of sensitivity and enhancement of specificity, likely concealed some small and isolated IR clusters, yielding fewer false positives for the presence of uniquely high or low IR findings, increasing false negatives (Figure 1 B,C). Although it seems reasonable to presume that low SES might contribute to elevated IR levels in sub-county areas assessed, the cross-sectional character of the SES and IR measures used in this study precludes determination of the temporal association between these characteristics. Findings from this study are unique to California, a state having counties that exceed the population size, geographic areas and demographic diversity of some US states and having distinctive geographic features that include the pacific coastline, a southern international border, mountain ranges and large, sparsely populated desert regions.

**CONCLUSION**

Cervical cancer is detectable during a treatable stage through
Pap testing and preventable through HPV vaccination, with inequitable distribution of prevention resources predicted by geography and SES. Findings presented here depict cervical cancer incident ratios and statistical patterns that reveal sub-county areas of California exhibiting higher than average occurrence of cervical cancer for 2006-2011. This inequity is predicted by lower than average SES in some geographic areas of California. Assuming that circumstances predicting the geographic patterns revealed in this study persist to today, it is reasonable to infer that this information could be used to diminish the occurrence of future cervical cancer cases in California.

Combined with local understanding of underserved population characteristics, this information could be used to target high risk segments of the California population through strategic placement of culturally and language sensitive health education and screening resources. Among these strategies, increased utilization of HPV vaccination [12] and Pap screening [56,57] according to current public health recommendations, targeted to sub-county geographic areas having higher than average IRs and in areas showing statistical clustering of cervical cancers cases, could be used to minimize risk for this preventable disease. It is our belief that incorporating these findings as part of current and planned cervical cancer prevention programs will enhance the success of those programs.

CONFLICT OF INTEREST AND FINANCIAL DISCLOSURE

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

REFERENCES

7. Austin RM, Zhao C. Type 1 and type 2 cervical carcinomas: some cervical cancers are more difficult to prevent with screening. Cytopathology. 2012; 23: 6-12.
14. Wilyman J. HPV vaccination programs have not been shown to be cost-effective in countries with comprehensive Pap screening and surgery. Infect Agent Cancer. 2013; 8: 21.


