Use of Spatial Epidemiology and Hot Spot Analysis to Target Women Eligible for Prenatal Women, Infants, and Children Services

| Thomas J. Stopka, PhD, MHS, Christopher Krawczyk, PhD, Pat Gradziel, PhD, RD, and Estella M. Geraghty, MD, MPH

The Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), a federally funded nutrition and health program, was established as a pilot program in 1972. WIC provides nutrition education, referrals, breastfeeding support, and nutritious supplemental food for low-to moderate-income families with nutritionally at-risk pregnant and breastfeeding women, infants, and children up to the 5th birthday. With an annual federal budget of $6.2 billion, the WIC program serves nearly 9 million participants each month through 1900 local agencies in state public health departments, Indian tribal organizations, and US territories. Additionally, the WIC Overseas Program provides services to Americans and dependents living overseas at US military bases. Nearly 30% of pregnant women participate in the WIC program nationally each year. In 2010, the WIC program served 62.6% of all who were eligible. WIC served 84.8% of eligible infants, 80.6% of eligible postpartum women (both breastfeeding and nonbreastfeeding), and 52.4% of eligible children aged 1 to 4 years. In California, WIC agencies provide services to nearly 1.45 million women, infants, and children each month at more than 625 sites statewide with an annual budget of approximately $1.2 billion.

Research focused on the prenatal benefits of the WIC program has not produced consistent conclusions. Some research has indicated that WIC participation is associated with improved birth outcomes and that provision of WIC services to pregnant women reduced low birth weights (<2500 g) and very low birth weights (<1500 g) by 30% and 54%, respectively, with substantial reductions in first-year medical costs for US infants. Moreover, prenatal WIC participation was found to reduce racial disparities in infant mortality rates. For example, the infant mortality rate for African Americans was significantly lower for WIC participants than for non-WIC participants. However, other studies have indicated that prenatal participation in WIC had minimal effects on adverse birth outcomes.

Not all women and children who are eligible to receive WIC services actually participate in the program. In 2007, only 59% of the total eligible US population participated in the program. Research has demonstrated that significantly higher rates of enrollment were observed for women who received prenatal care at health department- and community-sponsored clinics compared with hospital clinics and private physician offices. The number of prenatal visits was found to be positively associated with WIC participation rates among pregnant women, and the number of prenatal visits was found to improve birth outcomes. Previous participation in the WIC program was positively associated with early prenatal enrollment in WIC.

Spatial analytical methods and geographic information systems (GISs) have increasingly been used in public health, epidemiology, and nutrition research. Nutrition researchers have begun to recognize the importance of GIS and spatial analysis, particularly as they relate to measuring the role of the built environment and the food environment. GIS and nonstatistical (i.e., descriptive) mapping techniques have recently been used to explore disparities in access to fresh produce in low-income neighborhoods. We are unaware of any studies, however, that have used statistically based spatial analyses to assess WIC-eligible women who were not participating in WIC services.

Hot spot analysis is a statistically based method to assess geographic clustering. Specifically, hot spot analysis is used to pinpoint locations of statistically significant high- and low-value clusters of a phenomenon of interest by evaluating each feature (e.g., census tract).
within the context of neighboring features and against all features in the dataset. A feature with a high value may be a statistically significant hot spot if it is also surrounded by other features with high values, as opposed to simply being a data outlier. The local mean for a feature and its neighbors is compared proportionally with the global mean of all features (e.g., all census tracts in a state). When the observed local mean is much different than the global mean and that difference is too large to be the result of random chance, a statistically significant z score results and a hot spot cluster is detected. Recent studies using this technique have explored factors associated with West Nile virus incidence, tuberculosis transmission, patterns of sexually transmitted diseases, locations of health care services, and community-level overweight and obesity.

The purpose of our study was to identify the macro- and microlevel geographic regions of California that contain clusters of WIC-eligible nonparticipants. Findings were intended to provide the California WIC Program with an evidence base for funds allocation, to inform program decisions for a subset of counties and subcounty areas, and to strategically target specific populations and areas for WIC service enhancements (i.e., opening of new WIC centers).

METHODS

We linked data from the California Birth Statistical Master File (BSMF) for all women with a live birth in 2010 \( (n = 503,722) \) with program data from the California WIC Integrated Statewide Information System (WIC-ISIS). The BSMF consists of variables documented on the birth certificate by hospital staff. WIC-ISIS data are entered by WIC staff during each participant visit. We linked WIC-ISIS records to the BSMF database before linkage to WIC-ISIS. We geocoded residential addresses for mothers who gave birth in 2010 using the browser-based geocoder available through the California Environmental Health Investigations Branch (http://www.ehib.org/toolist.jsp). During the geocoding process, we compared each address with reference data for streets and addresses across California and ultimately assigned each a latitude and longitude to create a point on a map (similar to placing a pin on a map). We attained a 98% geocoding match rate (i.e., addresses matched a reference address in the geocoder). We then conducted a “point in polygon join” (i.e., reverse geocoding) within the GIS to assign a census tract number to each geocoded address point. We used census tract boundaries for 2000 because 2010 boundaries were not available until well after we completed our data linkage, reverse geocoding, and spatial analyses.

**Descriptive mapping.** We used descriptive GIS mapping techniques to assess the spatial distribution of WIC-eligible nonparticipants at the state, county, and subcounty levels. We explored the distribution of WIC-eligible nonparticipant counts, percentages of women who were WIC-eligible nonparticipants of all WIC-eligible women, and densities with thematic maps and corresponding data tables. These maps provided preliminary indications of the spatial distribution of WIC-eligible nonparticipants. In such thematic maps, however, it was not possible to reliably determine definitive patterns across the 7049 census tracts in the state. We conducted spatial analyses that invoked a number of statistical tests in our next phase of analysis to determine the location of statistically significant clusters of high densities of WIC-eligible nonparticipants.

**Spatial analysis.** We used a series of spatial analytical techniques to find clusters of WIC-eligible nonparticipants on the macro- (i.e., state) and microlevels (county).

**Step 1—Analysis of Variation in California Census Tract Areas**

In California, census tracts range in size from 0.02 to 7992.4 square miles, with a mean area of 22.4 square miles \( (SD = 173.79) \). This variation can cause analytic challenges in 2 ways. First, in descriptive mapping, one is visually biased by the bigger-sized tracts even though the measured outcome may be just as pronounced (or more so) in smaller tracts (census tracts are relatively homogeneous in population size and demographic makeup). Second, it is difficult to develop an appropriate neighborhood (local mean) size (i.e., spatial scale) with large variation in areas. To understand the degree of size variation in California’s 7049 census tracts, we calculated the square mile area of all census tracts in the state. We considered census tracts that possessed a square mile area greater than 1.5 standard deviations above the state mean census tract area outliers \( (n = 118) \) and temporarily removed them from the dataset. In addition, we also removed small census tracts that had been adjacent to the removed ones (and appeared as “islands”; \( n = 11 \)). This allowed us to begin to control for variability in census tract size in the next step of the analysis, considering census tracts of homogeneous geographic sizes.
Step 2—Determination of Spatial Scale, Part A
We began to determine the appropriate spatial scale for remaining California census tracts (n = 6920). Because our hot spot analysis examined each census tract in the context of its neighbors, we needed to develop an analytic sphere of influence around each census tract. We therefore calculated the average and maximum distance it would take to get from the geocentroid of each census tract to the geocentroid of its nearest 2 neighbors, deciding a priori that 3 neighboring census tracts with high or low WIC-eligible nonparticipant densities would constitute a cluster.

Step 3—Determination of Spatial Scale (the Moran I), Part B
To inform allocation of resources in the most focused area possible, we were interested in using the smallest possible distance (analytic sphere of influence) at which clustering of WIC-eligible nonparticipants was most intense, if clustering existed at all. If geographic analysis was to be useful, our outcome of interest (WIC-eligible nonparticipants per square mile) needed to exhibit spatial autocorrelation (based on Tobler’s first law of geography, which states that everything is related to everything else, but nearby things are more related than things that are far away28) to ensure that we found the shortest distance for which clustering (spatial autocorrelation) was intense. A common statistical measure of the degree to which a set of spatial features and their associated data values tended to be clustered or dispersed is the Moran I.21

We performed the Moran I test at multiple distances using the results from step 2. Our shortest distance tested was two thirds of the maximum distance it would take to reach the 2 nearest neighboring census tracts. Our experience has shown that using the actual maximum distance is often too conservative when trying to find the distance of most intense clustering at local geographic scales, because it is possible that all island census tracts were not removed or that larger census tracts continued to inflate the result of the distance-to-neighbors calculation. Subsequent Moran I tests were done incrementally, using half of the average distance to 2 neighbors found in step 2. Testing continued until the resulting z score peaked and was statistically significant (P<.05). The distance at that peak point (see Figure A, available as a supplement to the online version of this article at http://www.ajph.org) represented the smallest distance at which clustering of WIC-eligible nonparticipants was most intense and therefore represented an appropriate scale of analysis for subsequent hot spot analyses (step 5).

Step 4—Accounting for the Larger Polygons
Once we determined the distance that was the exemplar sphere of influence, or distance band, for the hot spot analysis, we next accounted for the large and island census tracts that had previously been left out of the analysis. To do this, we created a definition file called a spatial weights matrix that represented the spatial structure of our data and quantified the spatial relationships that existed among the features in the dataset. The structure of the matrix was an N × N table where N was the number of features in the dataset. Every feature in the dataset had 1 row and 1 column. We chose a binary strategy to quantify the relationships among data features in our dataset so that in the matrix a feature either was a neighbor (1) or was not (0). In this case, we included a feature as a neighbor if it was within the distance band calculated in step 3 or, for the previously excluded features, if it was 1 of the 2 nearest neighbors. With this strategy, every feature in the dataset was guaranteed 2 neighbors even if those neighbors did not fall within the local distance band parameter.

Step 5—Hot Spot Analysis
Finally, we conducted hot spot analyses to determine the locations of statistically significant clusters of census tracts with higher densities of WIC-eligible nonparticipants than would be expected given the mean density of WIC-eligible nonparticipants in the dataset.

We first conducted this 5-step geoprocessing approach to analyze the density of WIC-eligible nonparticipants for all 7049 census tracts in California to detect counties with clusters of WIC need relative to the state mean. Once we determined significant hot spot clusters and identified priority counties, we repeated the 5-step geoprocessing approach to identify more local clusters of census tracts relative to means of those priority counties.

RESULTS
Our results consist of 2 main components: (1) the step-by-step results that led to each subsequent stage in the hot spot analysis, drawing attention to the reproducibility and objectivity of the hot spot methodology, and (2) the locations in which we found hot or cold spots for WIC-eligible nonparticipants.

Descriptive Maps
Descriptive thematic maps portray the number, percentage, and density of WIC-eligible nonparticipants across California by counties and census tracts (Figure 1). Counts and percentages were categorized by quintiles. These maps indicated that counties in the Bay Area, the Central Valley, and the Los Angeles County Basin possessed large counts and densities of WIC-eligible nonparticipants (Figure 1a, 1c, 1d) and provided initial information about the burden of unmet WIC service needs. Figure 1b portrayed the percentage of WIC-eligible women who were eligible nonparticipants, providing initial indications of WIC service coverage.

Incremental Spatial Autocorrelation
The distance at which clustering of WIC-eligible nonparticipants was greatest, when considering all 7049 California census tracts, was 26 kilometers (16.2 miles; z score = 37.3; P<.001). The distance in county-specific analyses varied by county: Sacramento (distance = 5 km; z score = 9.1; P<.001 [see Figure A]), San Francisco (distance = 1.1 km; z score = 3.2; P=.001), Los Angeles (distance = 11 km; z score = 39.4; P<.001), and Fresno (distance = 9 km; z score = 10.6; P<.001).

Hot Spot Cluster Analysis
Results from hot spot analyses at the census tract level portray a detailed picture of the statistically significant clusters of WIC-eligible nonparticipants (Figure 2). Red census tracts denoted hot spot clusters with significantly higher densities of WIC-eligible nonparticipants than the mean density of eligible nonparticipants for all census tracts (P<.05). Yellow tracts represented census tracts that had densities of WIC-eligible nonparticipants that were not statistically different from the mean density of WIC-eligible nonparticipants in the state. Blue census tracts denoted cold spots,
Note. ENPs = eligible nonparticipants; WIC = Women, Infants, and Children.

**FIGURE 1**—Thematic maps of Special Supplemental Nutrition Program for Women, Infants and Children–eligible nonparticipants by (a) number at the county level, (b) percentage at the county level, (c) number at the census tract level, and (d) dot density at the census tract level: California, 2010.
or lower densities of WIC-eligible nonparticipants, that were significant at the $P<.05$ level.

In statewide analyses (58 counties), we found statistically significant hot spot clusters of census tracts with high densities of WIC-eligible nonparticipants in 21 counties: 7 counties in the Central Valley (Sacramento, Yolo, Placer, San Joaquin, Stanislaus, Madera, Fresno), 6 counties in the San Francisco Bay Area (Marin, San Francisco, Contra Costa, Alameda, San Mateo, Santa Clara), 3 Central Coast counties (San Benito, Monterey, San Luis Obispo), and 5 Southern California counties (Ventura, Los Angeles, Orange, San Bernardino, Riverside; $P<.05$). Significant cold spot clusters (blue shading), with low densities of WIC-eligible nonparticipants, existed in 37 counties, with notable clusters in the more rural regions of the state.

The San Francisco Bay Area, Los Angeles County, and the Central Valley, which encompassed large hot spots and a large number of WIC-eligible nonparticipants, merited closer attention. County-specific hot spot analyses revealed neighborhood-level clusters of census tracts with high densities of WIC-eligible nonparticipants, primarily in urban areas of the counties (Figure 3). We found local hot spot clusters of census tracts with high densities of WIC-eligible nonparticipants in the Downtown and North Highlands neighborhoods of Sacramento; the Tenderloin, Civic Center, South of Market, and Mission neighborhoods in San Francisco; the Downtown, Reseda, and Long Beach areas in Los Angeles; and the Central Valley.

**DISCUSSION**

GIS and spatial analyses have experienced considerable growth and popularity in public health in recent decades. The growing role of space and geography in understanding health issues and applying interventions underscores the fact that the field of spatial epidemiology, with ever improving hardware and software tools, is greatly needed.

In California, we used data visualization techniques with a GIS and conducted systematic and complex spatial analyses to determine the location of statistically significant clusters of census tracts with high densities of WIC-eligible nonparticipants in 2010. We used a 5-step geoprocessing approach, culminating in cluster-detecting hot spot analyses, which allowed us to first determine the California counties with greatest unmet need for WIC services. Then, in 4 counties, we performed local hot spot analyses to determine the opportunity neighborhoods with regard to clusters of WIC-eligible nonparticipants.

We detected clusters of WIC-eligible nonparticipants in the San Francisco Bay Area, along the Central Pacific Coast, in the Central Valley, and in the Los Angeles Basin. These clusters of high eligible nonparticipant densities appeared to parallel areas of high population density and large densities of families living in poverty (data not shown). County-level hot spot analyses in Sacramento, San Francisco, Fresno, and Los Angeles highlighted locations in which WIC directors could consider enhancing WIC services and outreach by increasing staffing, service hours, or the number of WIC sites available on the local level.

Conversely, statistically significant cold spots highlighted regions in which expanded WIC services may not have been needed and in
which WIC may have been adequately serving those communities’ needs.

The results of our spatial and statistical analyses have been presented to WIC directors, WIC staff, health care providers, and stakeholder groups through GIS maps, interactive online map presentations, and reports. The results have been used to guide resource allocation decisions, target outreach efforts, and help guide public health policy and program enhancement decisions. In Sacramento County, for instance, 1 local WIC agency used results from Sacramento hot spot cluster analyses, in concert with local data, to guide the opening of 2 new WIC centers.

Our study has several limitations. We focused on the location of residence at the time of birth in 2010 and, as a result, cluster analyses may not represent the reality of counties, cities, and local regions today. In our experience, however, we have not seen large changes in WIC service needs from one year to the next between 2008, 2009, and 2010, and our current findings are likely to represent needs for WIC services for the years immediately preceding and following 2010. Moreover, our methodology is highly reproducible and can be used annually to track change and improvements in programmatic services. Although our hot spot analyses allow us to determine locations in which WIC services may need to be modified to meet increasing or decreasing service needs, hot spot analyses do not allow us to determine the factors that are associated with such clusters of need. Multivariate logistic regression and spatial regression methods are needed to determine such associations and will be considered in subsequent analyses.

Finally, our calculation for WIC-eligible nonparticipants in California (n = 30 697) was based on Medi-Cal status (i.e., if women were on Medi-Cal, they were eligible for WIC services). Women may also have been eligible for WIC services if their income was lower than 185% of the federal poverty level, even if they were not receiving Medi-Cal benefits. We estimate that this could have expanded our population of WIC-eligible nonparticipants to approximately 51 300. Data for these larger eligible nonparticipant estimates were not
available for our analyses and will be considered for future research.

The federally funded WIC program has been around for nearly 4 decades and, although the program continues to provide important nutrition support and health education to millions of mothers, infants, and children across the country each year, the use of spatial data and complex spatial analyses foster fine-tuned assessments of program coverage and areas of refined service apportionment.

Our 5-step geoprocessing analysis that culminated in hot spot maps provided a rigorous and systematic method to determine the location of statistically significant clusters of WIC-eligible women. Use of these approaches in addition to traditional data visualization techniques (e.g., thematic maps) provides policymakers and program managers with an evidence base for important public health program and funding decisions. Similar analyses can be conducted for other public health programs to help assess the coverage and breadth of services in specified catchment areas that can facilitate targeting of public health services. During good budgetary times, hot spot analyses can point to counties, cities, and local neighborhoods in which services can be enhanced. During less favorable economic times, cold spot clusters can help inform policymakers and program directors to provide services in more efficient ways or relocate services to areas of higher need. Used in concert with statistical modeling approaches and spatial epidemiology, hot spot analyses provide additional tools for researchers, public health practitioners, and policymakers.

Contributors
T.J. Stopka was the lead author and conducted all analyses described. C. Krauzeck assisted with study design and implementation. P. Gradziel assisted with study design and implementation. E.M. Geraghty assisted with study design and spatial analysis. All of the authors provided input on interpretation of results and the writing of the article.

Acknowledgments
We thank the California WIC Program through the US Department of Agriculture for funding to conduct this study (universal ID 7C7A00CA7). We thank Catherine Camacho, director for MCAH, Center for Family Health, CDPH; Shabhir Ahmad, Barbara Longo, Michael Curtis, and Charlene Manning for support on this work and Katie Martin, Shylle Hob, and Archana Manul for contributions to earlier phases of analysis. We are grateful to Lauren Scott and Lauren Rosenheim Bennett for technical assistance with the hot spot cluster analysis methodology.

Human Participant Protection
This study was reviewed by the Committee for the Protection of Human Subjects, Health and Human Services Agency, State of California, and was deemed exempt.

References
28. Tobler W. A computer movie simulating urban growth for our analyses and will be considered for future research. The federally funded WIC program has been around for nearly 4 decades and, although the program continues to provide important nutrition support and health education to millions of mothers, infants, and children across the country each year, the use of spatial data and complex spatial analyses foster fine-tuned assessments of program coverage and areas of refined service apportionment.

Our 5-step geoprocessing analysis that culminated in hot spot maps provided a rigorous and systematic method to determine the location of statistically significant clusters of WIC-eligible women. Use of these approaches in addition to traditional data visualization techniques (e.g., thematic maps) provides policymakers and program managers with an evidence base for important public health program and funding decisions. Similar analyses can be conducted for other public health programs to help assess the coverage and breadth of services in specified catchment areas that can facilitate targeting of public health services. During good budgetary times, hot spot analyses can point to counties, cities, and local neighborhoods in which services can be enhanced. During less favorable economic times, cold spot clusters can help inform policymakers and program directors to provide services in more efficient ways or relocate services to areas of higher need. Used in concert with statistical modeling approaches and spatial epidemiology, hot spot analyses provide additional tools for researchers, public health practitioners, and policymakers.

About the Authors
At the time the study was conducted, Thomas J. Stopka and Christopher Krauzeck were with the Epidemiology, Evaluation and Data Operations Branch, Maternal, Child, and Adolescent Health Program (MCAH), California Department of Public Health (CDPH), Sacramento. Pat Gradziel is with the California Women, Infants and Children (WIC) Program, CDPH, Sacramento. Estella M. Geraghty is with the Department of Internal Medicine, School of Medicine, and the Graduate Group in Epidemiology, University of California, Davis, and CDPH, Sacramento. Correspondence should be sent to Thomas J. Stopka, PhD, MHS, Department of Public Health and Community Medicine, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111 (e-mail: thomas.stopka@tufts.edu). Reprints can be ordered at http://www.aph.htm.org by clicking the “Reprints” link.

This article was accepted October 19, 2013.