Title
Guidelines for the Mapping of Cancer Registry Data

Journal
Journal of Public Health Management and Practice, 19(3)

ISSN
1078-4659

Authors
Roberts, Eric M
Kumar, Bahar
Collins, Natalie
et al.

Publication Date
2013-05-01

DOI
10.1097/phh.0b013e318268aef1

Guidelines for the Mapping of Cancer Registry Data: Results From a Breast Cancer Expert Panel Study

Eric M. Roberts, MD, PhD; Bahar Kumar, MPH; Natalie Collins, MSW; Liang Guo, MS; Galatea King, MPH; Michelle Wong, MPH; Janice Barlow, BSN; Joyce Bichler, MSW; Linda Cady; Connie Engel, PhD; Debbie Garrett, BS, MHA; Marie Harrison; Adriana Morieko, MBA; Neena Murgai, PhD, MPH; Sora Park Tanjasiri, DrPH, MPH; Karen Pierce, JD

Context: Small area (eg, subcounty) cancer mapping is one of the analytic services most commonly requested of cancer registries and local public health agencies, and difficulties in providing it have been noted to undermine public confidence. Although a great many statistical protocols have been published to enable this practice, none of them are in common use to generate information for the general public. **Objectives:** To evaluate the utility of subcounty breast cancer mapping and articulate guidelines and a possible protocol for its implementation by cancer registries and local public health agencies. **Methods:** We convened an Expert Advisory Group of breast cancer stakeholders from around California to elicit values, priorities, and preferred characteristics of protocols for proactive subcounty breast cancer mapping. Upon formulating a protocol, we applied it to 9 years of data (2000-2008) describing invasive breast cancer in California for evaluation by the Expert Advisory Group. **Results:** Maps with subcounty resolution were seen to provide important information with a wide range of applications. Priorities included the avoidance of false-positive findings, scientific credibility, and the provision of information elucidating social and environmental characteristics. A protocol using Kulldorff's Scan Statistic along with postanalytic steps for refining results was elaborated; when applied to the data, 4 discrete regions with elevated rates of invasive breast cancer were identified and described. **Conclusions:** Expert Advisory Group priorities were readily translatable into a scientifically rigorous protocol that protected confidentiality and avoided statistically unstable rate estimates. The resulting maps enabled participants to visualize geographically defined populations falling within and crossing county boundaries. These findings support the enactment of policies for the routine and proactive analysis of breast cancer surveillance data to provide subcounty information.

KEY WORDS: breast cancer, community participatory research, disease mapping, cancer registries

J Public Health Management Practice, 2013, 19(3), E1–E10

Disease mapping is an effective means for conveying incidence and mortality information and is therefore one of the most commonly requested analytic services of local and state public health agencies.[1-4] Diseases such as breast cancer do not strike all members of society equally, and the complex interplay between vulnerability, stressors, environmental exposures, and

Author Affiliations: Public Health Institute, Richmond (Dr Roberts and Mss Kumar, Collins, King, and Wong and Mr Guo); Zero Breast Cancer, San Rafael (Ms Barlow); Breast Cancer Action, San Francisco (Ms Bichler); Between Women, Brawley (Ms Cady); Breast Cancer Fund, San Francisco (Dr Engel); California Health Collaborative, Fresno (Ms Garrett); Greenaction, San Francisco (Ms Harrison); Latinas Contra Cancer, San Jose (Ms Morieko), Alameda County Department of Public Health, Oakland (Dr Murgai); California State University at Fullerton (Dr Tanjasiri); and San Francisco County Department of Public Health (Dr Pierce), California.

This study was funded by the California Breast Cancer Research Program, University of California Office of the President (project 15UB-8405).

The authors declare no conflicts of interest.

Correspondence: Eric M. Roberts, MD, PhD, Public Health Institute, c/o Environmental Health Investigations Branch, 850 Marina Bay Pkwy, Bldg P, 3rd Floor, Richmond, CA 94804 (eric.roberts@phi.org).

DOI: 10.1097/PHH.0b013e318268aef1

genetics leads to patterns of disease that frequently echo disparities in resources and social privilege. Geographic patterns in disease reflect the multiple pathways and factors that contribute to population morbidity, most of which are only partially understood. When effective, disease mapping communicates these disparities in a manner useful to a broad array of advocates within the breast cancer community and can be valuable for any community seeking to understand its collective vulnerability and access to resources.[1,4-7]

Most geographic information provided by cancer registries and local public health agencies is confined to county-level reporting. Community advocates are typically frustrated by this limitation, given that the greatest needs for hypothesis generation relevant to environmental and social etiologies of cancers occur at the level of towns and neighborhoods. Although sweeping statements are difficult, these agencies tend to confine subcounty mapping activities to "cluster response" efforts, restricting data analyses solely to those requested by communities whose concerns have been subsequently assessed as valid by epidemiologists.[8-11]

Agencies following cluster response protocols therefore avoid analyzing cancer surveillance data beyond the county level until obliged to do so by pressure from parties external to the agency. Analyses are then explicitly confined to the area defined by those parties, which ensures that the foci of subcounty surveillance will always be chosen by those able to exert pressure on the agency rather than by scientific concerns. Several commentators have delineated the ways in which such reaction-oriented systems harm public trust and erode confidence in government agencies,[12-14] particularly because of community awareness that no data analyses would occur in the absence of outside pressure. Nationwide, state and local health departments receive between 1000 and 2000 inquiries regarding local elevations in cancer incidence (frequently referred to as "clusters") every year.[15]

A straightforward remedy to these problems would be the routine, proactive mapping of subcounty cancer incidence for entire states or regions so that areas with elevated risk could be defined and understood with scientific rigor prior to the onset of public concern. Commonly cited arguments against proactive mapping include concerns about confidentiality[16-19] and the domination of geographical patterns by random variation,[20] although both of these concerns have been addressed in the recent decades through statistical methods designed explicitly to meet them (Table 1). We posit that the nonutilization of these methods has 2 causes. One is that most public health agencies lack resources to address the statistical complexity of local variations in cancer risk, particularly when areas are small or sparsely populated.[28-30] The second cause is difficult to evaluate but no less cogent: there are actually too many options available, all associated with numerous decision points regarding their precise method of implementation. In this situation, the absence of a deliberative body having both authority and motivation to navigate these decision points is itself an obstacle to proactive cancer mapping.

The goal of the 2-year California Breast Cancer Mapping Project was to address the gap between these statistical advances and their implementation to generate information for communities and policy makers in California. Using an Expert Advisory Group (EAG) composed of collaborators with diverse backgrounds related to breast cancer and a multidisciplinary team of statisticians, geographic information specialists, and community health educators, we developed a broadly applicable protocol to help locate vulnerable communities, understand demographic risk factors, target prevention and intervention efforts, and help generate hypotheses regarding breast cancer etiology.

Methods

Although the focus of this project had many technical elements, we viewed our fundamental tasks to be the identification of EAG priorities, the translation of these priorities into technical protocols, the evaluation of protocol results, and the identification of essential communication and public health messaging concerns; our methodology was therefore devised according to a qualitative consensus panel model.[31] Therefore, considerations related to the composition of the EAG and its access to technical information were paramount, and we include them in the presentation of methods here.

EAG composition and support

The breast cancer advocacy community includes an array of public health care professionals, clinicians, and grass-roots organizers focusing on issues ranging from patient services to prevention and awareness to environmental action. Project staff included experts in biostatistics and epidemiology, medicine, community health education, and geographic information systems. Expert Advisory Group composition was geared to provide experts with knowledge and experience the staff did *not* have—specifically those with understandings of community information needs, agendas for social change, and strategies for health promotion.

The ultimate composition of the EAG is reflected in Table 2, and many of the participants appear as coauthors of this report. All participants were identified through a "purposive" sampling method commonly used for the identification of key informants.[32] In brief,

TABLE 1 ● Statistical Approaches Applicable to Subcounty Cancer Incidence Mapping and Cluster Detection (Partial List)

Model	Data Type	Comments	Key References
Conventional choropleth	Area	Possibly the only approach used on a routine basis outside of academia Assumes statistical independence of disease risk in geographic areas regardless of proximity or adjacency Least statistical power of all approaches	Waller and Gotway (2004)[21]
Scan Statistic	Area or point	Along with hierarchical Bayesian modeling (later), most discussed and evaluated approach in academic literature Able to evaluate statistical significance of detected "clusters" while adjusting for multiple testing	Kulldorff (1997)[22]
Hierarchical Bayesian modeling	Area	Along with the Scan Statistic (mentioned earlier), most discussed and evaluated approach in academic literature In most common formulation (BYM) posits spatial structure in data as one of several variables, the distributions of which are described through Gibbs and/or Metropolis sampling	Besag et al (1991)[23]
Flexible Scan Statistic	Area	Statistically similar to the Scan Statistic but searches for irregularly shaped "clusters"	Tango and Takahashi (2005)[24]
SPATCLUS	Area or point	Evaluates interpoint distance between cases or area centroids Requires "buffer" language for communication, although these may be of irregular shape	Dematteï et al (2006)[25]
Density estimation	Area or point	Assumes a priori degree of spatial structure ("kernel size") and formulates event densities at points determined by this structure	Rushton and Lolonis (1996)[26]
Generalized additive models	Point	Uses nonparametric terms (loess or spline) to depict spatial variation Can determine ideal degree of spatial structure using deviance-based criteria such as Akaike's information criterion	Webster et al (2006)[27]

TABLE 2 ● Expert Advisory Group Composition

Affiliation	Participant Background
Alameda County Department of Public Health	Public health epidemiology
Bayview-Hunters Point Health and Environmental Assessment Task Force	Environmental justice
Between Women	Breast cancer patient support
Breast Cancer Action	Advocacy
Breast Cancer Fund	Environmental health advocacy
Breast Cancer Task Force, American Cancer Society of California	Medical provider
California State University, Fullerton	Academia
California Health Collaborative	Breast cancer early detection
Greenaction	Environmental justice
Latinas Contra Cancer	Community-based organization
UCSF Helen Diller Cancer Center	Academia
Zero Breast Cancer	Community-based organization

Abbreviation: UCSF, University of California San Francisco.

staff engaged in open-ended phone interviews during which they described the project and solicited peer recommendations for participants. Staff also sought participants representing a diversity of professional, cultural, and economic backgrounds within California. Although the location of the staff in Northern California meant that this region had predominant representation within the EAG for logistical reasons, participants included those from other regions such as the state's Central Valley, Imperial Valley, and Greater Los Angeles. Although this composition did not include representation of all California communities, staff and participants agreed that this degree of diversity among the EAG was one of its strengths.

Communication strategies

The levels of EAG participants' understanding of cancer surveillance and biostatistics ranged from novice to expert. Valid discussion required the establishment of a common language for communication and the use of concrete examples and graphics to express otherwise abstract statistical concepts. For example, a series of maps depicting idealized cancer "clusters" annually over several years was used showing that the measured incidence and boundaries of an area of concern underwent random fluctuations over time. Such a series of graphics was used to facilitate discussion of the statistical concept of stochastic variation, although such wording was not explicitly used. A further requirement for effective group dialogue was the recognition by the group that all participants approached the project with different areas of expertise. Therefore, mutual respect for the diversity of backgrounds represented by the participants was explicitly recognized, and contributions enabled by the multiplicity of perspectives were actively sought and articulated in discussions.

Articulation of priorities

Although EAG participants were eager to use the forum provided by the project to share experiences and develop strategies among their peers, they recognized (and appreciated) that their needs were diverse. Within that context, staff developed 3 themes around which questions ("probes") were used to provide structure and overall focus to meetings.[31] These themes included the (1) utility (or nonutility) of subcounty breast cancer mapping; (2) methodological questions (ie, how it should be done); and (3) contextual issues, such as the utility of supporting information such as social and environmental risk factors for breast cancer.

Staff developed several strategies both to encourage concreteness in discussion and to enable participants having nontechnical backgrounds to weigh in on topics that might otherwise be prohibitively abstract. For theme 1, this included the provision of maps of fictional communities for which vulnerable populations, labor and economic structures, and health resources were described in detail. Discussion included ways in which hypothetical elevations in breast cancer might be interpreted by each community.

For theme 2, this included a browser-based interactive mapping application displaying how simulated cancer "clusters" would look using various statistical approaches. Participants were able to manipulate the sensitivity and specificity of these approaches and compare output with "true clusters" created using simulated data. Participants were encouraged to explore the mapping approaches by using the application both singly and in groups and to record their responses by using written forms designed to solicit open-ended feedback as well as group discussion.

The statistical approaches used in the application corresponded to those used in journalism and public health communication as well as more sophisticated techniques that were well established in the literature. These included the calculation of risk ratios (observed/expected) often used in the popular press, the calculation of tract-specific *P* values under the assumption of complete statistical independence classically used by public health care professionals, and 2 of the algorithms developed by statisticians (spatial Scan Statistic and hierarchical Bayesian modeling) listed in Table 1.

Expert Advisory Group responses were recorded and consolidated in a variety of ways. For the interactive mapping exercise, participants handwrote answers to both closed- and open-ended questions. Two facilitators were dedicated to note taking for all in-person meetings, and phone meetings were recorded; typed versions of notes were subsequently circulated to participants for commentary and corrections.

Operationalization of priorities as statistical protocols

Even considering the extensive discussion and capacity building in which the EAG was engaged, we recognized that full understanding of any mapping protocol could not be achieved without carrying the protocol through to completion using actual cancer surveillance data. Biostatistics staff were assigned the task of translating the priorities articulated by the EAG into protocols that would use one or more of the approaches mentioned earlier (or others if necessary). For this work, staff paid particular attention to EAG feedback relevant to themes of statistical power, accessibility of the format of results, potential for various types of confounding, and limitations imposed by data quality. Their reasoning was then presented to the EAG for validation and discussion before using the protocols for statewide analysis.

Statewide analysis

Because measures of in situ breast cancer would be confounded by screening rates, EAG and staff chose invasive breast cancer incidence as the outcome of interest. To this end, we used annual counts of new diagnoses of invasive breast cancer among California women by (year 2000) census tract and age for the years 2000 to 2008. Although any results would involve aggregations of data necessary for the protection of confidentiality, the analyses themselves required the handling of sensitive data; therefore, approval had been sought and obtained from the Committees for the Protection of Human Subjects of the California Department of Health Services and the Public Health Institute.

Data were obtained from the California Cancer Registry, having been collected and managed according to standards set by the Centers for Disease Control and Prevention's National Program of Cancer Registries and National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program. Cases were defined as new diagnoses of breast cancer (SEER diagnostic code 26000) among females for which the stage was not recorded as "in situ." Cases lacking confirmation through microscopy or solely reported through autopsy, death certificate, or an outpatient center were excluded. All records included a residence address for the time of diagnosis that was geocoded by a commercial geocoder as an exact street match; failing an exact match, the centroid of the ZIP+5 boundary was used.

Denominator data were drawn from the US Census counts from 2000 to 2010. Since age-specific counts of women for 2000 census-based tracts were not available for 2010, these data were calculated from 2010 census-based tracts through reapportionment according to the population weights supplied by the Bureau of the Census. Denominators for between-census years were then generated through linear interpolation.

Findings

EAG priorities

The following are thematic summaries based on multiple EAG meetings, including individual feedback forms. All of these were articulated in different ways during these meetings but appeared with sufficient consistency for staff and EAG participants to consider them representative and reliable.

- The greatest value of subcounty breast cancer maps lies in their potential for conveying surveillance data in a manner that is both intuitive and concise. Although much of the information they are likely to show will be consistent with known demographic risk factors in the general population, they add value by helping frame discussions of patient services, interventions, and causes. Because of issues of latency and population mobility, the EAG considered maps for other types of cancer (eg, pediatric cancers) to be more likely to generate hypotheses regarding environmental determinants than those for breast cancer. It was felt, however, that this did not detract from the maps' overall utility. Furthermore, stakeholders concerned with local environmental hazards potentially associated with breast cancer were able to quickly assess whether their concerns were corroborated by a discrete geographic "cluster"; if not, they were able to set this question aside and pursue other relevant sources of scientific information.
- As false-positive elevations are misleading and distract communities from otherwise useful action, statistical specificity was crucial in the selection of any mapping protocol. Several participants with extensive experience working with public health agencies had been aware of the danger of false-positive detections, yet they voiced surprise that methods existed for excluding them. The fact that alternative methods for avoiding false-positive findings were so rarely mentioned among agency personnel and other public health care professionals was itself a focus of a discussion.
- The EAG's conception of "specificity" contained a nuance not anticipated by statistical staff. False-positive elevations located near the boundary of true areas of elevation were not considered to be as problematic as those occurring far from true areas of elevation. Although findings in the latter category were clearly misleading, those in the former were interpreted as imprecision regarding the boundaries of true elevations. Although this imprecision raised important communication issues for any protocol, the difference in interpretation—"there is a true area of elevation here but we don't know exactly where its boundaries lie" versus "there is an apparent area of elevation entirely due to random chance"—was felt to be crucial.
- Besides specificity, credibility among the scientific community was considered the major criterion for method choice. Among the EAG were seasoned activists who were aware of the pitfalls of laboriously assembling data only to learn after the fact that standards of scientific acceptability had not been met. Several participants asked probing questions regarding the degree to which analytic methods—such as spatial Bayesian modeling and Scan Statistic—were considered controversial versus established among scientists.
- The potential utility of maps for communication with the public was considered contingent on the provision of contextual information, including discussions of data sources, interpretation and its limitations, additional resources for when maps did not answer specific questions, and guidance for requesting data and information from public agencies overall.
- Despite the excitement generated by the subcounty maps, conventional presentations of data (including county-level maps) still had their place. Part of this is due to the fact that resources are commonly distributed at the municipal and county levels, so data may be required that match this geography. Also, the flexibility of subcounty mapping by necessity requires a loss of statistical power; therefore, areas with predetermined geography (such as counties) might show statistically significant elevations in breast cancer rates with conventional methods but not appear on proactively generated subcounty maps.

Operationalization of priorities as statistical algorithms

As a starting point, staff considered which protocols enabled users to maintain the highest levels of specificity (using the nuanced EAG definition) while preserving levels of sensitivity participants found meaningful on the basis of their responses to the interactive browser-based exercise. Although they had been aware that traditional calculations of risk ratios and independent *P* values were inadequate to this task, staff further realized that the unique algorithm used in the SaTScan software offered specific advantages in this respect. The spatial Scan Statistic algorithm, which involves data-specific Monte Carlo simulations, focuses on statewide rather than location-specific probability of type I error.[33] In other words, the translation of Scan Statistic *P* values (eg, $P = .001$) as "1 false-positive cluster anywhere in the state among 1000 sets of similar analyses" rather than "1 in 1000 locations having a false-positive cluster," matched the stated EAG requirement for avoidance of misleading false-positive detections with a degree of certainty conventional and even spatial Bayesian methods could not match.

Staff noted that the selection of the Scan Statistic as an analytic method did not constitute a complete protocol on its own. For example, the software designed by Kulldorff and colleagues (SaTScan) requires users to make an array of analytic and reporting choices. Staff felt that they were able to make these decisions on the basis of the general principles arising from EAG discussions; these choices are presented in the upper section of Table 3. Although the logic behind most of the SaTScan parameter settings is self-explanatory, that of the Maximum Spatial Cluster Size requires discussion. As noted in Table 3, communication requirements dictated that only nonoverlapping circular buffers would be reported. When the maximum buffer radius is high, small numbers of large "clusters" covering entire regions of the state are reported; on the basis of the EAG discussions, staff concluded that such results would have limited utility. When the maximum buffer radius is low, few areas of the state have sufficient population density such that any elevation of reasonable severity can be detected. Staff reasoned that, at some midpoint between these 2 extremes, there would be a setting conveying the maximum amount of information reflected in the highest number of unique areas identified within the 9 years of data. Exploration confirmed this view, with the greatest number of individual "clusters"

TABLE 3 ● Operationalization of Expert Advisory Group Priorities as Analytic Protocols

Decision Point/Problem	Solution	Comments
SaTScan settings		
Probability model type	Poisson	Dictated by data format
Coordinates	Latitude/longitude	Used 2000 census-based tract centroids
Type of analysis	Purely spatial	Based on EAG interest
Scan for areas with . . .	High rates	Based on EAG interest
Monte Carlo replications	9999	Enables calculation of *P* down to 10^{-4}
Maximum spatial cluster size	30-km radius	Maximal resolution based on data exploration (see text)
Spatial window shape	Circular	Elliptical shapes require a priori specification of noncompactness penalty, etc
Criteria for reporting secondary clusters	No geographical overlap	Alternatives yield large numbers of areas of concern, thus limiting communication utility of results
Postanalytic processing		
Nonlinear population growth can lead to 1- to 2-y "clusters" based on denominator error	Suppress "clusters" appearing for $\leq$3 y	Transient elevations in breast cancer are of questionable validity anyway; due to population mobility over long lag periods, environmental exposures from single sources are unlikely to result in geographically discrete "clusters"
Require nonchanging cluster boundaries so that trends can be observed over time	Define areas of concern as the total of tracts included in any "cluster" not suppressed as earlier	Small numbers of areas of concern facilitates communication and discussion
Require understanding of incidence changes (or persistence) over time	Using earlier definition, construct time-series plots	Frequently, these plots suggest ongoing elevations even in years for which no "cluster" is detected
Findings only meaningful with contextual data	Using earlier definition, characterize areas of concern on the basis of available sociodemographic and clinical variables	Based on EAG interest

Abbreviation: EAG, Expert Advisory Group.

occurring when the 30-km radius limit was used (data not shown).

Staff developed protocols for further winnowing Scan Statistic results postanalysis (lower section of Table 3). For example, the 2000-2008 period witnessed rapid changes in the geography of housing stock that were likely to result in nonlinear population growth in many locations; the resulting errors in population estimates were therefore expected to give rise to transient, spurious elevations during the period between censuses. To guard against distraction by findings based on these or other artifacts (eg, a 1-time increase in breast cancer screening resulting in earlier detection for a subpopulation), staff adopted the criterion that areas must contain a detected cluster during at least 3 of the 9 years examined. These decisions were shared with and approved by the EAG in subsequent discussions.

Statewide analysis

Data were analyzed 1 year at a time for the 2000-2008 period; for this period, SaTScan identified a total of 30 unique "clusters." Of these, 6 were excluded for being consistent with known (nonlinear) shifts in population and/or appearing only transiently. The remaining "clusters" were readily grouped by location into 4 distinct areas (Figure, left). These included southern Orange County (177 tracts, appearing in 3 of the 9 years), southern San Francisco Bay (264 tracts, 4 years), northern San Francisco Bay (360 tracts, 8 years), and western Los Angeles plus eastern Ventura Counties (699 tracts, all 9 years). Time-series plots of each area of concern (Figure, right) suggest that, with few exceptions, invasive breast cancer rates in each area during detected years were not demonstrably different from those during nondetected years, which reinforces the impression that the patterns noted are consistent over time. The magnitude of each increase was generally between 10% and 20% for any given year.

For reasons of space, the reader is referred to the full report (www.californiabreastcancermapping.org) for analysis of demographic and clinical variables for each area of concern; in that report, we have endeavored to follow the guidelines produced by the EAG regarding the provision of findings and contextual information in an accessible manner. Notably, the findings tended to reinforce previous understandings of demographic risks (eg, elevations for women of European descent). Similarly, analysis of patient data for each of the areas suggested that slightly fewer patients relied on public sources of payment for their health care than the state average, with the exception of western Los Angeles/eastern Ventura, for which slightly more did so. These analyses are presented in their entirety in the online report.

EAG conclusions

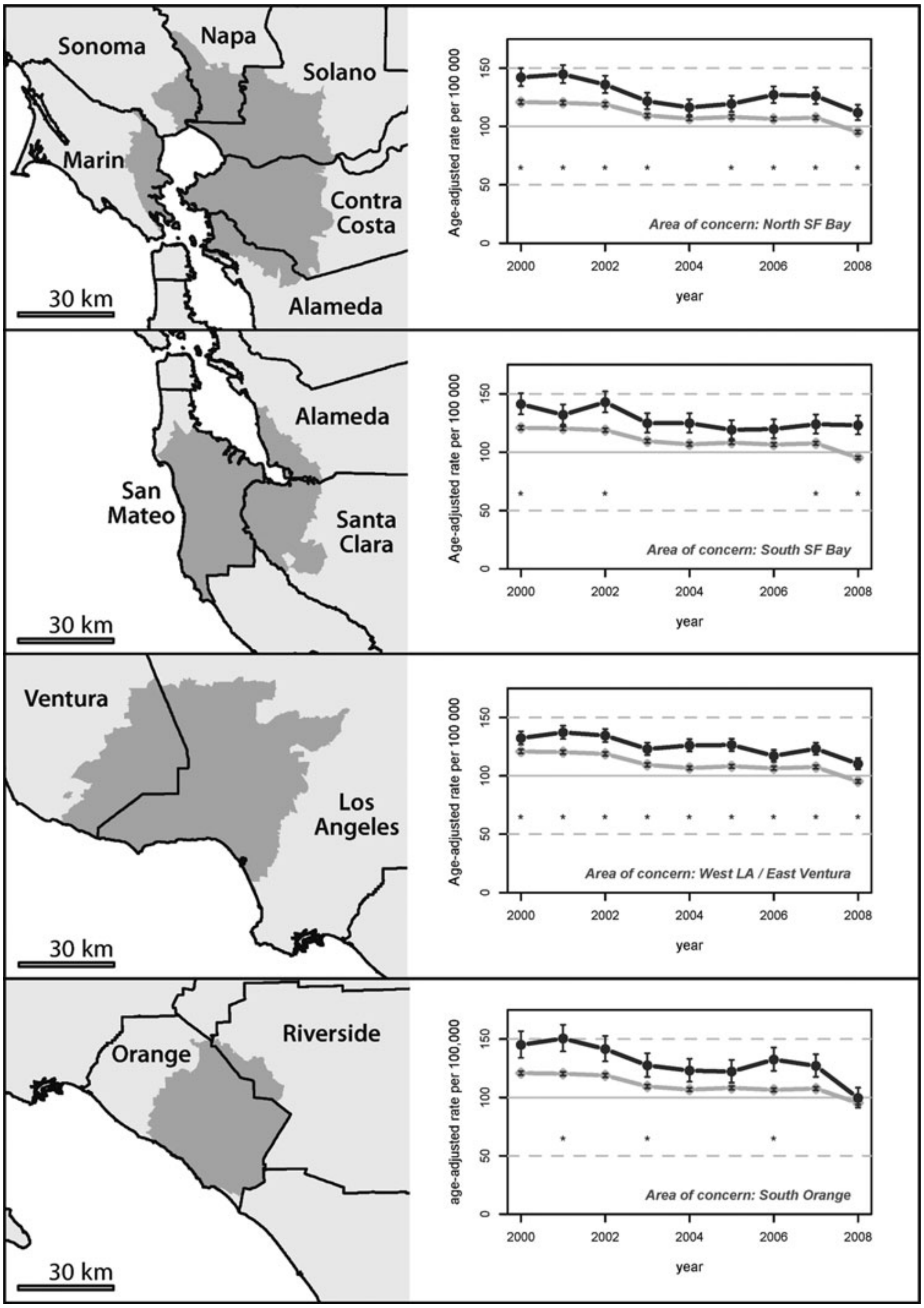
The EAG considered the subcounty maps produced by their protocol valuable not as a replacement to other modes of presenting breast cancer surveillance data but rather as a supplement to them. For example, funding and interventions are often conceived as county-level initiatives, so there will always be a need for county-level figures describing incidence and prevalence that already exist. At the same time, it was clear that the addition of the information coming from the maps lent to discussions a greater sophistication than had previously been possible. Neither Los Angeles nor Ventura County has been noted to have elevated rates of breast cancer relative to the state as a whole, so the ability to consider an area of concern that crossed the boundary of the counties represented a substantial improvement. The attention that could be drawn to the southern portion of Orange County or subsets of the San Francisco Bay region was considered similarly useful.

Several EAG participants voiced disappointment that information was not generated for rural portions of the state, and there was concern that this might be due to limitations in statistical power. However, we knew from our extensive simulations for the browser-based application that elevations of the size and severity of those found in urban areas could also be found in rural areas. In this respect, the findings are consistent with county-level analyses, which generally demonstrate that counties in rural regions of California have invasive breast cancer rates that are not persistently elevated over time. It is notable, however, that census tracts are physically larger in rural areas; therefore, our 30-km radius limit may make elevations in rural areas more difficult to detect.

California is one of the most populous states in the United States, with several counties containing larger and more diverse communities than many smaller states. Although the EAG considered the protocol useful overall, they predicted that it might be most successfully applied to a number of contiguous states in combination if attempted elsewhere—for example, all of New England or a grouping of 2 or more Midwestern states.

The question of the applicability of the protocol to the surveillance of other types of cancer was also frequently discussed by the EAG. Since the incidence of most (but not all) cancer types are orders of magnitude smaller than that of breast cancer, questions of sensitivity and specificity would need to be revisited, and adjustments to the protocol might be warranted. As noted earlier, the long latency period between hypothesized exposures to environmental carcinogens and the development of breast cancer precludes the use of this approach to investigate questions of environmental

FIGURE ● Age-Adjusted Rates of Invasive Breast Cancer (With 95% Confidence Intervals) Among Females by Area of Concern, 2000-2008[a]



[a]Left, Map labels refer to counties that include census tracts of interest. Right, Rates in areas of concern are represented by black circles; statewide rates are represented by gray diamonds; and asterisks indicate years for which Scan Statistic indicated an elevation somewhere within each area of concern. LA indicates Los Angeles; SF, San Francisco.

causes in breast cancer. Similar hypotheses for other cancer types (particularly among children) may not involve long latency periods; therefore, the implications and messaging associated with the dissemination of analytic results for these cancer would need to be modified to address the possibility that environmental factors could play a role in any findings. In a similar fashion, known demographic patterns in breast cancer do not hold for most other types, so descriptions of populations with highest risk would also differ.

Finally, there was concern among the EAG that the subcounty maps generated were rarely (if ever) produced or disseminated by cancer registries or local public health agencies in general practice. The long-standing nonadoption of methods such as the Scan Statistic by public health agencies was difficult for the EAG to rationalize, particularly in light of their known adherence to standards of scientific rigor (a quality sought by all stakeholders, including those represented in the EAG). Throughout the exercise, virtually no risks appeared related to patient confidentiality, all calculations included numbers well above those required for the calculation of stable rate estimates, and elevations that appeared likely to arise from denominator error were easy to detect and discount. Participants recognized that there was ambivalence on the part of public agencies for the utilization of this or similar techniques, but the logical justification for this ambivalence (beyond the characteristic slowness of institutional change in general) was difficult to understand.

Conclusions

Surveillance—defined as the routine and ongoing collection, analysis, and dissemination of public health information—is considered a cornerstone of public health practice.[34] Although mapped representations of the subcounty geography of cancer incidence—including breast cancer incidence—are commonly sought by diverse audiences, the routine production of such information is generally not practiced by public agencies. Through systematic engagement and discussion of priorities by a diverse expert panel of breast cancer advocates in California, we elaborated and tested a protocol suitable for the detection and mapping of areas with elevated incidence of invasive breast cancer that could be implemented on a routine, proactive basis. We found these results to be readily communicable as a publicly accessible report (www.californiabreastcancermapping.org) under further guidelines articulated by the EAG.

The protocol yielded information that—while strictly adhering to all protocols protecting patient confidentiality and using noncontroversial statistical methods long established in the scientific community—was considered to represent a valuable improvement for the understanding and communication of geographic patterns of breast cancer. Because of these findings, the EAG strongly supports the enactment of policies for the routine and proactive analysis of breast cancer surveillance data in this manner, along with the exploration of the protocol's suitability for other types of cancer.

REFERENCES

1. Bell B, Hoskins R, Pickle L, Wartenberg D. Current practices in spatial analysis of cancer data: mapping health statistics to inform policy makers and the public. *Int J Health Geogr.* 2006;5:49.
2. Roberts E, English P, Wong M, Wolff C, Falade M. Continuous local rate modeling for communication in public health: a practical approach. *J Public Health Manag Pract.* 2008;14(6):562-568.
3. Roberts EM, English P, Wong M, et al. Progress in pediatric asthma surveillance II: geospatial patterns of asthma in Alameda County, California. *Prev Chronic Dis.* 2006;3(3):A92.
4. Hébert J, Daguise V, Hurley D, et al. Mapping cancer mortality-to-incidence ratios to illustrate racial and sex disparities in a high-risk population. *Cancer.* 2009;115(11):2539-2552.
5. Beyer K, Rushton G. Mapping cancer for community engagement. *Prev Chronic Dis.* 2009;6(1):1-8.
6. Driedger S, Kothari A, Morrison J, Sawada M, Crighton E, Graham I. Correction: using participatory design to develop (public) health decision support systems through GIS. *Int J Health Geogr.* 2007;6:53.
7. Heitgerd J, Dent A, Holt J, et al. Community health status indicators: adding a geospatial component. *Prev Chronic Dis.* 2008;5(3):1-5.
8. Aldrich T, Sinks T. Things to know and do about cancer clusters. *Cancer Invest.* 2002;20(5/6):810-816.
9. Greenberg M, Wartenberg D. Communicating to an alarmed community about cancer clusters: a fifty state survey. *J Community Health.* 1991;16(2):71-82.
10. Kingsley B, Schmeichel K, Rubin C. An update on cancer cluster activities at the Centers for Disease Control and Prevention. *Environ Health Perspect.* 2006;115:165-171.
11. Thun M, Sinks T. Understanding cancer clusters. *CA Cancer J Clin.* 2004;54:273-280.
12. Siegrist M, Cvetkovich G, Gutscher H. Shared values, social trust, and the perception of geographic cancer clusters. *Risk Anal.* 2001;21(6):1047-1053.
13. Tajik M, Gottlieb K, Lowndes N, Stewardt G. Environmental policies, politics, and community risk perception: case study of community contamination in Casper, Wyoming. *New Solut.* 2007;17(4):345-361.
14. Trumbo C, McComas K, Besley J. Individual- and community-level effects on risk perception in cancer cluster investigations. *Risk Anal.* 2008;28(1):161-178.
15. Trumbo C. Public requests for cancer cluster investigations: a survey of state health departments. *Am J Public Health.* 2000;90(8):1300-1302.
16. Brownstein J, Cassa C, Mandi K. No place to hide—reverse identification of patients from published maps. *N Engl J Med.* 2006;355(16):1741-1742.
17. Cox L. Protecting confidentiality in small population health and environmental statistics. *Stat Med.* 1996;15:1895-1905.
18. Olson L, Grannis S, Mandl K. Privacy protection versus cluster detection in spatial epidemiology. *Am J Public Health.* 2006;96(11):2002-2008.
19. VanWey L, Rindfuss R, Gutmann M, Entwisle B, Balk D. Confidentiality and spatially explicit data: concerns and challenges. *Proc Natl Acad Sci U S A.* 2005;102(43):15337-15342.
20. Bellec S, Hémon D, Clavel J. Answering cluster investigation requests: the value of simple simulations and statistical tools. *Eur J Epidemiol.* 2005;20:663-671.

21. Waller LA, Gotway CA. Applied Spatial Statistics for Public Health Data. Hoboken, NJ: John Wiley & Sons; 2004:78-83.
22. Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods*. 1997;26:1481-1496.
23. Besag J, York J, Mollié A. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Statist Math*. 1991;43(1):1-59.
24. Tango T, Takahashi K. A Flexibly Shaped Spatial Scan Statistic for Detecting Clusters. *Int J Health Geogr*. 2005;4:11.
25. Dematteï C, Molinari N, Daurès J. SPATCLUS: An R package for arbitrarily shaped multiple spatial cluster detection for case event data. *Comput Methods Programs Biomed*. 2006;84(1):42-49.
26. Rushton G, Lolonis P. Exploratory spatial analysis of birth defect rates in an urban population. *Stat Med*. 1996;15(7-9):717-726.
27. Webster T, Viera V, Weinberg J, Aschengrau A. Method for mapping population-based case-control studies: an application using generalized additive models. *Int J Health Geogr*. 2006;5:26.
28. Boulton M, Abellera J, Lemmings J, Robinson L. Assessment of epidemiologic capacity in state and territorial health departments—United States, 2004. *MMWR Morb Mortal Wkly Rep.* 2005;54(18):457-459.
29. Hopfer S, Chadwick A, Parrott R, Ghetian C, Lengerich E. Assessment of training needs and preferences for geographic information systems (GIS) mapping in state comprehensive cancer-control programs. *Health Promot Pract.* 2009;10(4):579-587.
30. Juzych N, Resnick B, Streeter R, et al. Adequacy of state capacity to address noncommunicable disease clusters in the era of environmental public health tracking. *Am J Public Health.* 2007;97:S163-S169.
31. Green J, Thorogood N. Group interviews. In: *Qualitative Methods for Health Research*. 2nd ed. Los Angeles, CA: Sage; 2009:123-146.
32. Liamputtong P. Theoretical frameworks and sampling in qualitative research. In: *Qualitative Research Methods*. 3rd ed. South Melbourne, Australia: Oxford University Press; 2009: 2-18.
33. Information Management Services IncKulldorff M. *SaTScan^{TM} v8.0* [computer software for the spatial and space-time scan statistics]. http://www.satscan.org/. Published 2009. Accessed November 30, 2009.
34. Brookmeyer R, Stroup D. *Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance*. New York, NY: Oxford University Press; 2004.