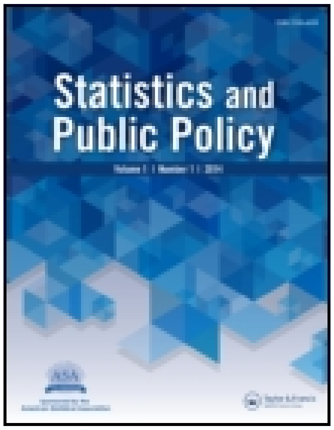

Statistics and Public Policy
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uspp20

Discussion: Statistical Cluster Detection, Epidemiologic Interpretation, and Public Health Policy

Lance A. Waller

Published online: 17 Apr 2015.

To cite this article: Lance A. Waller (2015) Discussion: Statistical Cluster Detection, Epidemiologic Interpretation, and Public Health Policy, Statistics and Public Policy, 2:1, 1-8, DOI: 10.1080/2330443X.2015.1026621

To link to this article: http://dx.doi.org/10.1080/2330443X.2015.1026621

Discussion: Statistical Cluster Detection, Epidemiologic Interpretation, and Public Health Policy

Lance A. Waller

We briefly review five articles exploring spatiotemporal clustering of pediatric cancer cases in the state of Florida for the years 2000–2010. We review the general motivating question of interest (Are there clusters in the data?) and compare approaches with specific attention to the statistical quantification of this question, the epidemiologic insight gained about disease patterns, and potential policy responses to the collective results.

KEY WORDS: Cancer clusters; Epidemiology; Spatial statistics.

1. INTRODUCTION

Many thanks to the authors of the five articles investigating statistical clusters of pediatric cancer in Florida for the years 2000–2010. Having multiple groups analyze the same dataset with different methods offers an opportunity for deeper insight into the construction, application, and especially interpretation of statistical methods to detect geographic areas and time periods with unexpectedly high incidence rates. It also offers an opportunity to review epidemiologic insights revealed by the analyses and potential policy responses to address public concerns. Past examples of multiple articles analyzing shared disease clustering datasets (e.g., Alexander and Boyle 1996) often focus on a comparison between methods to see which methods perform the best by some criterion, say, the greatest statistical power to detect clusters of various sorts. While interesting, such an approach tends to build on a yes/no statistical goal of detecting a cluster (or not) then focuses on which method is "better" in making this distinction. As we'll see in the sections below, an alternative approach is to think beyond a dichotomous "cluster found/not found" outcome and to look carefully at where, when, and how methods agree as well as where, when, and how they differ to provide a fuller description of the underlying patterns of disease and population at risk to aid the transition between statistical results, epidemiologic insight, and policy decisions.

In the present case, the five author groups apply different approaches to the data, each with the same general goal (to detect spatio-temporal clusters) but differing mathematical and statistical implementations and sensitivities. The key general *statistical* question is whether the observed patterns of reported cases appear consistent with a probability model of case assignment to individuals at risk *in the absence of clustering* (often referred to as "assignment by chance"), or whether there are local collections of cases in time and space that appear statistically inconsistent with such a model (patterns "unlikely to have arisen by chance"). While the general question is the same, it is addressed in slightly different ways within each of the five articles and these differences reveal specific and interesting aspects of the underlying disease patterns contained within the data.

There is a considerable literature regarding disease cluster detection and its perceived and actual utility in epidemiology and public health. On the statistical side, Lawson (2006) and Waller and Gotway (2004, chaps. 6 and 7) provided broad overviews of statistical methods and probability models (of "no clustering"), and Tango (2010) provided a comprehensive catalogue of statistical hypothesis testing-based approaches for detecting clusters and clustering. The five articles here include application and modification of some familiar methods (scan statistics, classification, and hierarchical Bayesian modeling) as well as some ideas new to disease cluster detection (wombling and machine learning).

Based on my own work in the area (Waller and Gotway 2004; Waller 2009), I find it helpful to discuss the five approaches with respect to the following four specific questions:

- What general question do we want to answer?
- What data are available?
- What specific questions does each method answer with the available data?
- What do the answered questions reveal about our motivating general question?

Within each of these questions, I focus on three conceptual dimensions: the *statistical* aspects (are the identified clusters statistically unusual, based on an underlying model of "no clustering"?), the *epidemiological* aspects (do the observed patterns suggest the presence of unknown or suspected sources of increased risk, and, if so, do these patterns offer insight on potential causes of disease?), and *public policy* aspects (given the observed data and analytic results, is policy action called for?).

2. WHAT QUESTIONS DO WE WANT TO ANSWER?

As noted above, in the most general sense, each method seeks to answer the same question: "Are there unusual clusters of disease in the data?" The modifier "unusual" is important, since local aggregations of cases due to local aggregation of people at risk or of known risk factors are not of primary interest. For instance, a high number of observed cases in an area with a high number of individuals at risk may not indicate a local increase in individual *risk*, rather simply a larger local *number* at risk (e.g., many cancer cases at a retirement community need not

Lance A. Waller is Professor and Chair, Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322 (E-mail: *lwaller@emory.edu*).

Published with license by American Statistical Association
© Lance A. Waller
Statistics and Public Policy
e1026621 (8 pages)
DOI: 10.1080/2330443X.2015.1026621

suggest pollution, since the elderly are more prone to cancer). Most methods of cluster detection will define the number of local cases expected in the absence of clustering based on the size and composition of the local at-risk population with respect to known risk factors (e.g., age, race, and sex of the at risk population) and seek to detect local aggregations in space and time that are above and beyond those anticipated by local aggregations of such risk factors.

2.1 Statistical Questions

From a statistical perspective, cluster detection presents a very interesting challenge. Can we define the distribution of values we would expect to see in each location in the absence of clustering, and accurately and reliably detect deviations from this distribution? The approaches used by the authors build on this general motivating question using ideas from hypothesis testing (Amin et al. 2014), model-based (posterior) estimation (Heaton 2014; Lawson and Rotejanaprasert 2014; Zhang, Lim, and Maiti 2014), and machine learning (Wang and Rodriguez 2014). Again, all authors are interested in the statistical identification and evaluation of local aggregations of cases in space and time to see if there are any aggregations (clusters) that are inconsistent with known factors driving incidence. That said, each article takes a slightly different approach to operationalize these questions based on the available data and the methods used. We explore these differences in some detail below.

2.2 Epidemiologic Questions

Speaking very broadly, the general motivation in epidemiology is the accurate and reliable identification and quantification of risk factors associated with the onset, progression, treatment, and prevention of disease. With respect to cluster studies, the motivating epidemiologic question expands on the statistical question above: *Does the observed pattern of disease reveal new insights regarding potential causes of local increases in disease risk?* While both the statistical and epidemiologic questions are based on the observed pattern of cases, the epidemiologic question not only assesses whether the observed pattern is unusual (with respect to known risk factors) but also whether any observed anomalies reveal specific epidemiologic insight into local factors driving risk. That is, does a statistical excess of cases necessarily suggest a true underlying local increase in risk due to some local factor? As discussed by Amin et al. (2014), this question is often specified even further to focus on potential local environmental exposures, that is: *Does the observed pattern of disease reveal new insights regarding potential local environmental exposures causing local increases in disease risk?*

Epidemiology works best when one has a clearly defined population at risk with well-measured exposures, demographics, and disease outcomes. Cluster studies represent only one type of approach in the epidemiologist's toolbox, but a general cluster detection study is neither the first nor the preferred approach among epidemiologists for obtaining the best estimates of risk and risk factors due to several reasons. Cluster studies rarely have a well-defined study population with individual-level risk information (here, we start with all children in the state and all childhood cancers in the state of Florida, as reported by the Census and a pediatric tumor registry, respectively) or well-defined individual level exposures (here we use location as a proxy for unmeasured shared exposures to environmental factors). Thun and Sinks (2004) provided a very readable overview of cancer cluster interpretation from the epidemiologic perspective, noting that cancer cluster studies work best when they involve a rare cancer in a well-defined group of people with a specific, prolonged, high-intensity exposure to the same industrial or medical carcinogen. Well-known historical examples (detailed in Thun and Sinks 2004) include clusters of mesothelioma and lung cancer among asbestos workers and osteosarcoma among radium dial painters. In each case, the exposures were high, similar, and concentrated in a particular group of individuals, allowing detailed examination of medical outcomes, cumulative exposures, and similarities and differences between individuals as well as comparison to individuals without the exposure under investigation. Thun and Sinks (2004) noted that such clearly defined cluster studies are few and far between and note that, in contrast, most cluster investigations include a large population of individuals with different baseline risk factors (e.g., age, race, and sex), different occupational exposures, different lifestyle factors (e.g., level of exercise, diet), etc., considerably complicating any clear identification of an underlying shared exposure within suspected clusters of cases.

The epidemiologic literature contains many assessments of the role and potential value of cluster detection studies in revealing new causal insights for disease risk. Many of these are critical (e.g., Rothman 1990; Coory and Jordan 2013), noting that, with the exception of the very few remarkable instances mentioned above, rarely has detection of a cluster by itself directly revealed novel epidemiologic insight into unanticipated risk factors. As mentioned above, a primary epidemiologic concern is the heterogeneity of exposures experienced by individuals within any suspected cluster (particularly if that cluster contains a large number of individuals).

A second epidemiologic concern regards the accuracy of small area disease rate estimates. The accuracy of local estimated rates is driven by the accuracy of the local case count and the accuracy of the local count of individuals at risk. Inaccuracies in an observed increase in a local area's risk estimate could be due to a single case mistakenly assigned to an incorrect address or to an underestimated number of individuals at risk. Thun and Sinks (2004) cited two examples of the latter situation where initial suspiciously high cancer rates in small local areas within particular race groups were found to be due to underestimation of inter-Census migration of individuals in that race group into the area. When these projections of migration were corrected in the next decennial Census, the observed numbers of cases were found to be consistent with the numbers expected based on the corrected population demographics. Geographically small urban census areas with high population density, high rental property concentration, and high population mobility are particularly sensitive to such migration adjustments, a point we return to in discussion of the results from the five articles below.

The third basic epidemiological concern links the epidemiologic and statistical perspectives, that is, are the observed high rates statistically significant, that is, higher than we would expect to occur by chance? The use of statistical significance (*p*-values) is often a controversial topic in the epidemiologic literature, and accurately assessing significance of a suspected cluster is a challenging statistical issue (addressed in different ways by the authors below), leading some to call provocatively for an end to the use of statistical significance in cluster studies altogether (see recent article by Coory and Jordan 2013, and associated discussions).

In addition to data quality and statistical concerns, some epidemiologic critics question the logic of searching for potential causes of disease *after* identifying an area of higher-than-expected incidence, pointing out that *some* area will have the highest local risk and it can be difficult to determine whether the highest rate is unusually high due to some local cause of increased risk. These critics point out that only focusing on the highest local rates for epidemiologic follow-up studies can lead to the so-called "Texas sharpshooter" effect: shooting the side of the barn then painting a bulls-eye around the bullet hole and claiming to have hit the mark. This is particularly an issue with *perceived* clusters reported by the general public, for example, a community group notes several cases in their neighborhood and reports this to the county or state health department who then assesses statistical significance. Such an approach can miss high rates in areas without vigilant community members or physicians. While this situation differs from the current articles where a statewide registry is examined, it is a very common setting for cluster reports and an issue addressed repeatedly in the epidemiologic literature.

As an example of the epidemiologic view of the value of investigating perceived cancer clusters, Goodman et al. (2012) provided a thorough review of 428 published cluster investigations evaluating 567 cancers categories (some investigations include multiple cancer types). These authors note that only 72 (13%) of these cancer categories confirmed an unusual increase in cancer incidence (newly diagnosed cases). Three of these were linked (with varying degrees of certainty) to hypothesized exposures and only one investigation out of the original 428 revealed a clear cause for the increase (12 of 19 cases of pleural cancer shared a concentrated occupational exposure to asbestos in a shipyard). This review again illustrates that cancer cluster investigations very rarely lead to novel epidemiologic insight regarding unanticipated local risk factors causing disease.

While not all of these epidemiologic concerns apply to the analyses in the five analyses here, the concerns do provide an important context for expectations of epidemiologic response to statistical cluster detection and our discussions of policy implications below.

2.3 Policy Questions

Taking things to the next level, the motivating policy question is: *Does the observed pattern of disease suggest necessary or recommended policy responses?* Potential policy responses may involve more detailed follow up including, for example, additional data collection, exposure surveys, or in-depth investigation of case histories within the cluster. As mentioned above, state and local health departments and public health agencies regularly respond to cluster reports from the general public. In spite of the epidemiologic criticism of cluster studies above, Bender et al. (1990) and Neutra (1990) noted there remains an obligation to be "responsibly responsive" to concerns from the public regarding potential local sources of disease risk.

Thun and Sinks (2004) provided an overview of U.S.-based national agencies with protocols for cluster response some of which were recently updated, for example, that of the U.S. Centers for Disease Control and Prevention and the Council of State and Territorial Epidemiologists (Abrams et al. 2013). In their overview, Thun and Sinks (2004) cited literature reporting over 1000 reports of perceived clusters per year, indicating a sense of the policy demand for response. While the number of reports exceeds that that can be followed up with dedicated, detailed epidemiologic studies, Bender et al. (1990) and Thun and Sinks (2004) noted that, typically, a responsive and effective response is not based on a detailed *de novo* epidemiologic study, but rather based on education (e.g., clarifying that different cancers typically have different etiologies and may not be due to the same risk factors), assessments of local concentrations of demographic risk factors (age, race, sex) associated with the reported cluster, and an assessment of the distribution of numbers of cases expected (in the absence of clustering) given the local demographics. That is, most responses include a review and assessment of features in the data to clarify patterns rather than the initiation of a new, follow-up epidemiologic study, a point to keep in mind in our discussion below.

3. WHAT DATA ARE AVAILABLE?

All five articles use the same sets of cancer and population data, namely pediatric cancer case reports from the Florida Association for Pediatric Tumor Programs (FAPTP), a consortium of diagnosis and treatment centers that consolidates their data for incorporation into the Florida Cancer Data System (FCDS), part of the National Program of Cancer Registries of the U.S. Centers for Disease Control and Prevention. The FCDS is certified for timeliness and completeness by the North American Association of Central Cancer Registries (NAACCR), and incorporates the FAPTP data. The FCDS follows up on any discrepancies, omissions, or inconsistencies between the two. While the FCDS represents the highest quality data regarding cancer diagnosis and treatment in the state, the FCDS and FAPTP work closely to consolidate all data on pediatric cancers, and the FAPTP data are used in many epidemiologic studies in the state (as referenced in Amin et al. 2014). Amin et al. (2014) were clear that, in their analysis, they focused on the three most common categories of pediatric cancers (leukemias, lymphomas, and central nervous system cancers). Lawson and Rotejanaprasert (2014) specified brain cancer but the other authors are not clear whether they also limit attention to specific cancers or all pediatric cancers reported to the FAPTP. Since different cancers typically have different etiologies and since Amin et al. (2014) found slightly different clusters for the different cancers, it is epidemiologically important to clarify which cancers are included in each of the statistical analyses to best interpret results and define any appropriate policy response.

The data are summarized into ZIP code tabulation areas (ZCTAs), geographic areas defined by the U.S. Census Bureau to provide a link between Census geography (blocks, block groups, and tracts) and U.S. Postal Service ZIP code areas. Medical records often include a billing address (including a ZIP code) and ZCTAs were developed in part due to the large demand for a way to link such billing records (and other address lists) to Census demographic summaries without undergoing a complex geocoding exercise based on each individual address. It is important to recognize that ZCTAs are *close* to areas defining ZIP code delivery areas, but that they are not identical to ZIP code areas (which can and do change at any time based on postal delivery needs). Some ZIP codes in urban areas correspond to single buildings or large businesses, some rural ZIP codes meander along delivery routes. ZCTAs are constructed from complete census regions (e.g., blocks) combined together so that a large majority of residents of a ZCTA share the same ZIP code, but ZIP codes can (and often do) follow different boundaries than Census blocks (see articles by Grubesic and Matisziw 2006 and Beyer, Schultz, and Ruston 2008 for more detailed discussion). In the present analyses, it is important to note that the ZCTA data does not fully cover the state of Florida—a particular gap (due to low population density in the Everglades) is seen in the southernmost tip in the maps in figures 1 and 2 in Heaton (2014), figures 1 and 2 in Wang and Rodriguez (2014), figures 1–11 in Lawson and Rotejanaprasert (2014), and figures 1 and 3 in Zhang, Lim, and Maiti (2014). It is important to appreciate that ZCTAs provide a close approximation to link registry-based disease cases to census demographics defining the population at risk, but that it is still an approximation and may merit closer investigation to fully understand potential clusters, particularly those based on small local numbers of individuals at risk.

It is also important to note that both the FAPTP and the Census report race characteristics but in different categories, so it is somewhat difficult to precisely match the racial demographics of the cases with the demographic profile of the at-risk population. More specifically, the U.S. Census captures self-reported race and ethnicity through individual responses to two separate questions. For the 2000 Census, responders chose one or more self-identification classifications from the list of: "White," "African American/Black/Negro," "American Indian/Alaskan Native," "Asian Indian," "Japanese," "Native Hawaiian," "Chinese," "Korean," "Guamanian/Chamorro," "Filipino," "Vietnamese," "Samoan," "Other Pacific Islander," or "some other race." Ethnicity is determined by answering a question "Is this person Spanish/Hispanic/Latino?" from the following options "No, not Spanish/Hispanic/Latino", "Yes, Mexican, Mexican American, Chicano," "Yes, Puerto Rican," "Yes, Cuban," "Yes, other Spanish/Hispanic/Latino" (Grieco and Cassidy 2001). In contrast, the FAPTP reports race in only three categories "White," "Black," and "Other," with over 95% of the population falling into the first two categories (Ren et al. 2012). Race is reported from the medical record (or death certificate for mortality data) and may not necessarily be self-report but rather an assessment assigned by a physician, nurse, or other professional when filling in the medical or death record.

The last data feature to appreciate relates to the statistical challenge of small area estimation, here, estimating low risks/rates of a rare event from small local population sizes. We mentioned above the concern regarding accuracy of the local case count and population count data, particularly for between-Census estimations. There is also a tension between statistical and geographic precision: as we increase geographic precision by using smaller areas we decrease statistical precision by estimating rates based on fewer and fewer events. Some ZCTAs contain larger numbers of individuals at risk, allowing better precision in estimation, but some likely have less than one case expected. All statistical approaches applied in the five articles seek to combine or borrow information between ZCTAs to improve local estimates either through combining data across neighboring ZCTAs (the set of potential clusters evaluated in Amin et al. 2014 and the clusters of equal risk in Zhang, Lim, and Maiti 2014), or through spatial smoothing/correlations between ZCTAs (Heaton 2014; Lawson and Rotejanaprasert 2014; Wang and Rodriguez 2014). All authors note that their at-risk population values are from the Census and it appears they limit these to age groups consistent with the pediatric at-risk population, but to aid in epidemiologic interpretation of the statistical results, it is important to clarify this detail.

4. WHAT SPECIFIC QUESTIONS DOES EACH METHOD ANSWER WITH THE AVAILABLE DATA?

We next examine each article in terms of what specific details relating to the general question of interest are addressed by the methods applied. We begin with statistical details, followed by the epidemiologic and policy implications of the statistical findings.

4.1 Statistical Questions: Are There Statistically Significant Local Increases in Risk?

Amin et al. (2014) applied space–time scan statistics to identify the most unusual collection(s) of cases from a large set of potential clusters, then evaluate whether the most unusual collection is more unusual than one would expect under random allocation of cases among individuals in the at-risk population. More specifically, the scan statistic approach considers a large number of potential clusters, here collections of ZCTAs whose centroids fall in circles of radii ranging from the minimum intercentroid distance to an upper bound defined by a proportion of the entire state population. The approach assigns a score to each potential cluster under consideration to measure how unusual the rate observed inside the potential cluster is compared to the rate outside the potential cluster, then evaluates whether the extreme (highest) scores are unusually extreme (recall that *some* potential cluster will have the highest score, even if there is no clustering) compared to the distribution of extreme scores one would expect under uniform allocation of the observed number of cases among the at-risk population, adjusted for age and sex. The approach follows a hypothesis-testing framework where the null distribution of interest is the distribution of the maximum score(s) arising from the most unusual clusters arising under a model of no clustering. The score assigned to each cluster is based on a likelihood ratio *statistic*, but it is important to note that statistical significance is not evaluated via a likelihood ratio *test*. Rather, each statistic provides a measure of "unusualness" for a potential cluster and the key element of a scan statistic is the comparison of the observed maximum value of "unusualness" to the distribution of maximum values arising under the null hypothesis of no clustering, not evaluating each likelihood ratio statistic against its own null distribution. Amin et al. (2014) examined childhood brain/central nervous system cancers, childhood lymphomas, childhood leukemias, and the set of these three classes of pediatric cancers. For each, they answer the statistical question of interest: *Is the observed score associated with the most likely cluster higher than we would expect if there were no clustering?* Amin et al. (2014) found statistical evidence for higher than anticipated rates of brain/central nervous system cancer incidence in an area covering Miami (fig. 1 of Amin et al. 2014) and extending across south Florida, statistical evidence for higher than anticipated rates of leukemia incidence in a nearby area, also including Miami but extending west (fig. 3 of Amin et al. 2014), and statistical evidence of higher than anticipated rates of lymphoma, also in Miami (fig. 4 of Amin et al. 2014). An analysis of the three classes of cancer combined reveals statistical evidence for higher than anticipated incidence across South Florida (as might be expected from the three disease-specific analyses) and also a second area including Gainesville, Jacksonville, and St. Augustine. Both of these clusters are geographically quite large and the southern cluster spans across the ZCTA gap associated with the Everglades noted in the discussion of the data sources above (a feature that is difficult to ascertain in figs. 1, 3, and 4 of Amin et al. 2014).

Rather than evaluate the most unusual from a set of potential clusters, Heaton (2014) examined spatial and temporal boundaries between ZCTAs for each year, to identify the boundaries across which we observe the largest changes in rates. Heaton's approach is to identify the boundaries spanning the largest local changes between ZCTAs rather than identifying the ZCTAs with the highest rates. Conceptually, one can think of a series of ZCTA maps for each year where the expected number of cases under no clustering is based on age, race, and sex subpopulations. The statistical question addressed is: *Which boundaries suggest the greatest changes in rates and are these changes greater than we would expect if there were no clustering, after adjusting for age, race, and sex?* Note that this is similar, but not identical to the question addressed by Amin et al. (2014). Heaton (2014) specifically adjusted for race after preliminary analysis indicated different geographic concentrations for both cases and the at-risk population. Figure 2 of Heaton (2014) indicates areas of high incidence (presumably for all pediatric cancers), and that these areas are different for different race groups. In particular, Heaton's output suggests that the higher incidence west of the Everglades and near Orlando occurs primarily in the "Other" race category in the FAPTP data (not African American, Caucasian, or unknown). The results also indicate large rate changes as one moves into ZCTAs in Miami across all race groups. Heaton's approach identifies more localized patterns than does Amin et al.'s (2014) scan statistic approach, primarily because the search for high-change boundaries includes many more potential clusters of different shapes than does the set of circular clusters considered by the scan statistic approach. Heaton (2014) also examined potential effects of income and poverty variables from the Census data adding some further descriptions of the observed patterns. Associations with income variables vary by reported case race variables and suggest potentially important interactions between racial and economic demographics associated with the diagnosis of pediatric cancers. Finally, Heaton (2014) noted a suggestion of an increase in the baseline incidence rate occurring in 2005. It would be interesting to see whether this jump represents an overall increase in risk or a change in reporting, as the statistical analysis alone does not reveal potential reasons for the change. The interactions between high incidence areas, race, income, poverty, and reporting provide additional interesting insights not readily apparent in the scan statistic results and provide data-based recommendations for further investigation.

Lawson and Rotejanapraset (2014) and Zhang, Lim, and Maiti (2014) both developed hierarchical Bayesian models based on Poisson regression models with random effects to allow borrowing of strength across small areas to stabilize local estimates. However, the two approaches are customized to address different specific dimensions of the shared general question of interest. The goal of Lawson and Rotejanapraset (2014) was to improve ZCTA-specific estimates of risk by expanding traditional disease mapping models to account for the rarity of pediatric cancers, which results in more observed zero counts than one might typically expect in a Poisson distribution. More specifically, while the Poisson distribution is often used to model count data, particularly for counts of events among small areas, the distribution assumes a strong linkage between expectation and variance and is not flexible for modeling count data with extremely small expectations. Lawson and Rotejanapraset (2014) extended the model to a zero inflated Poisson formulation allowing a parameter to model the excess zeros and their results indicate a better fit to the data. Their Bayesian implementation provides posterior distributions for the local ZCTA rates and the authors summarize results by reporting ZCTA-specific posterior probabilities of exceeding certain risk thresholds. The question operationalized by this approach is: *Which ZCTAs have high estimated risk values AND high posterior probabilities of exceeding particular risk thresholds?* In the article by Lawson and Rotejanapraset (2014), the results in figures 3 (estimated relative risks) and 7 (estimated probability of these relative risks exceeding 1) indicate high posterior probability for higher risk in a particular ZCTA west of the Everglades (also highlighted in Heaton's fig. 2(c) for "Other race"). High individual ZCTA estimates also appear in the Miami area (although they are more difficult to see in the map due to the small geographic size of these high population density urban ZCTAs) and in other isolated areas across the state. We see some concordance of Lawson and Rotejanapraset's high smoothed ZCTA rates with high posterior probabilities and areas identified by Amin et al. (2014) and Heaton (2014), but we note that, unlike the previous two applications, Lawson and Rotejanapraset do not seek to identify consolidated areas of high risk, rather they seek to improve risk estimates and inference for each of the ZCTAs.

In contrast, Zhang, Lim, and Maiti (2014) used a hierarchical model with the express purpose of categorizing ZCTAs into geographically contiguous groups sharing a common risk level. They operationalize our general question of interest into: *Can we classify all ZCTAs into a limited number of geographic areas, each with a different risk level?* Rather than identify only high-risk areas, Zhang et al.'s classification approach identifies both the number of apparent risk categories and assigns ZCTAs to these categories, requiring geographic contiguity among members of the same category. Zhang, Lim, and Maiti (2014) noted that this approach of spatial classification has been used in the disease mapping literature, but we should also note that the classification goal of "clustering" observations into a set of categories is slightly different than the "cluster detection" goal of identifying isolated anomalies used more explicitly in the approaches of Amin et al. (2014) and Heaton (2014). Zhang, Lim, and Maiti's (2014) initial descriptive analysis in their figure 1, and their classification results (figs. 2 and 3) reveal that the number and location of their categories vary by year. Figure 1 in the article by Zhang, Lim, and Maiti (2014) provides further insight into the isolated ZCTA west of the Everglades noting that it reports a very high incidence rate for the year 2000 only. When put in context with the other results, the pattern suggests this may be due to a very small number of leukemia or brain/central nervous system cancer cases (from Amin et al. 2014) reported in a very low population density area in the "Other" race category (from Heaton 2014). The classifications by year in Zhang et al.'s figure 3 give little indication of any spatially consistent local increases over the study period, but do provide an overview of how pediatric cancer incidence rates vary across the state in space and time. Zhang, Lim, and Maiti (2014) proposed a metric for assessing the probability of a classification being a "hotspot" by assessing the posterior probability that samples of the rate within the "hotspot" exceed those in surrounding areas. While statistically interesting, the measure seems to be overly sensitive (79% of applications to simulated data under a model of no clustering detected a significant cluster) and appears to be more of a measure of the probability that the highest rate is higher than its neighbors, not an assessment of whether it is higher than we would expect the highest rate to be under a null hypothesis of no clustering. The latter sort of assessment is better accomplished by the scan statistic used by Amin et al., 2014. This is a subtle but important difference in the specific question addressed by elements of the proposed method.

Wang and Rodriguez (2014) brought a machine learning perspective to the investigation that in some respects focuses the classification question addressed by Zhang, Lim, and Maiti (2014) a bit more on anomaly detection. Specifically, Wang and Rodriguez (2014) used penalized likelihood estimation to identify ZCTAs with observed rates near the statewide rate, those significantly above this rate, and those significantly below this rate. The penalty induces smoothing so that small deviations around the background rate are ignored while more sizable differences remain. The specific question addressed then becomes: *Which ZCTAs exhibit incidence rates that are appreciably different from the statewide background rate?* Like Heaton (2014), Zhang, Lim, and Maiti (2014) adjusted for age, race, and sex in determining the statewide background rate. Zhang, Lim, and Maiti (2014) noted some similarity between typical disease mapping spatial smoothing priors and their spatial penalty. As a technical aside, it is interesting to observe that Zhang et al.'s penalized likelihood is mathematically identical to the formulation of a conditionally independent likelihood with a *pairwise difference prior* defined in Besag, York, and Mollié's (1991) landmark disease mapping article (see also Besag et al. 1995, eq. 3.3). The approach differs from the standard conditional autoregressive (CAR) prior by focusing on an absolute value difference in neighboring values (rather than the traditional squared difference for a CAR prior) to avoid oversmoothing large jumps in local rates between contiguous regions. Besag et al. (1995) noted that this approach is a stochastic version of a median filter used in image analysis yielding an interesting historical sequence of a penalized smoother from image analysis motivating a spatial smoothing prior for Bayesian disease mapping, which in turn motivates Wang and Rodriguez' spatial smoothing penalized likelihood. Historical interest aside, the connection suggests an even closer link between spatial penalized likelihood and disease mapping models than indicated by Zhang, Lim, and Maiti (2014), revealing that the penalized likelihood approach and a hierarchical Bayes disease mapping approach (with this particular prior) both base inference on an identical function. The difference lies only in the next steps of implementation. A penalized likelihood perspective views this as a function to be optimized with variances, etc., based on asymptotic arguments. The Bayesian approach sees this as a function to be sampled from to provide the full posterior distribution, and such an approach has been applied to disease mapping problems in the past (Besag, York, and Mollié 1991; Best et al. 1999). Zhang et al.'s figure 1 reveals the estimated relative risks resulting from their approach. The results indicate high estimated incidence in the Miami area and in a band where the panhandle meets the rest of the state (in addition to other, smaller areas). The same general pattern appears when adjusting for age, race, and sex. Heaton's (2014) race-specific figure 2(b) results provide further insight and reveal this later area primarily corresponds to an area of increased risk among Caucasians.

Taken together, the results suggest statistical evidence for areas of increased incidence within the study period, but we see subtle differences between the locations identified by the different methods. Clarifying the statistical question addressed by the various approaches provides some insight for potential differences and some slight variations in the applications also reveal additional features in the detected patterns, for example, the type of cancer (leukemia or brain/central nervous system cancer, as suggested by the results of Amin et al. 2014), the racial composition ("Other," as suggested by the results of Heaton 2014), and the year (2000, as suggested by the results of Zhang, Lim, and Maiti 2014) of the cluster west of the Everglades. In other words, the methods do not completely agree on the precise location, boundaries, and make-up of the detected clusters, but, in some ways, these differences, coupled with the detailed questions addressed, provide insight into potential epidemiologic and policy conclusions that would be missing in the application of only one method.

4.2 Epidemiologic Questions: Do We Find Factors Causing Disease?

Statistically, we have evidence of unusual increased risk in areas in South Florida near Miami, and a few other isolated locations around the state. We next focus on epidemiologic implications and interpretations of these clusters. We will consider three epidemiologic aspects related to the estimated rates and risks: which outcomes are associated with detected clusters, are these consistent and sustained within space and time, and what are the impact of demographic aspects of the at-risk group?

What outcome? We first note that while all five articles started with the same data, they each made subtle (and appropriate) adjustments to the data in their analysis. Amin et al. (2014) focused attention on three particular pediatric cancer classifications and illustrated that the cluster detection methods are somewhat sensitive to the outcome under study (consistent with potentially different etiologies between cancer types). Lawson and Rotejanaprasert (2014) focused on brain cancer, and the others seem to have included all pediatric cancers listed in the FAPTP (but this is not clearly stated). While the three cancer types used by Amin et al. are the most prevalent among childhood cancers, it would be interesting to see what, if any, changes result from focusing the other methods on the brain/central nervous system cancers, leukemias, and lymphomas.

Do we find consistent, sustained clusters over space and time? Some of the author groups specifically considered time in their analysis, revealing slightly different spatial patterns in different years. Amin et al. (2014), Heaton (2014), Wang and Rodriguez (2014), and Zhang, Lim, and Maiti (2014) all illustrated some changes in the detected patterns over years, mostly revealing that some of the small outlying clusters are confined to particular years. The results of Zhang, Lim, and Maiti (2014) revealed that the cluster west of the Everglades was confined to only the year 2000. From the epidemiologic perspective, the lack of spatial and temporal consistency for a long period speaks against a long-term, consistent local risk factor driving the observed patterns.

Is there an impact of age, race, and sex in the at-risk population? Some author groups adjusted for age and sex, others for age, race, and sex. Heaton's maps illustrate the importance of race in the observed outcomes, and, as mentioned above, this seems particularly important in interpreting the high local rate west of the Everglades.

Do the cluster results provide insight regarding potential causal factors? Amin et al. (2014) provided an overview of some past studies of potential environmental risks (e.g., drinking water, air pollution, and proximity to nuclear power facilities). Amin et al. (2014) noted that most of the past epidemiologic studies had not found significant associations between such exposures and disease risk in local populations. The clusters identified by Amin et al. (2014) and by Zhang, Lim, and Maiti (2014) are geographically quite large and unlikely to provide clear links between local concentrations of particular individual exposures to individual local risks, based on the general points raised by Thun and Sinks (2004) and the past epidemiologic experiences summarized in Goodman et al. (2012).

4.3 Policy Questions: What Do We Do?

The five author groups find some consistent results: there seem to be local areas where the observed cancer rate is statistically significantly higher than we would expect and the different methods tend to identify a few common areas. As noted above, the clusters themselves are not identical, but they do overlap and by looking across the five sets of results we observe the following features in the results:

- All methods identify statistical increases in the Miami area and in an area just west of the Everglades.
- The Miami cluster includes an urban area with geographically small ZCTAs and, based on past experience reported in Thun and Sinks (2014), a first step in follow up is to assess the accuracy of both the case counts and the Census-based population at risk. The case counts come from the FAPTP with specified quality control. The Census counts, particularly for years between decennial Censuses, merit a close look to see if population migration projections are accurate, particularly for large urban areas like Miami.
- The cluster west of the Everglades is included in the large south Florida leukemia and brain cancer/central nervous system clusters identified by Amin et al. (2014) but we lose the distinction of the cluster from the Miami area and the geographic gap in ZCTAs due to the Everglades in their figures 3 and 5. Heaton's (2014) results reveal that the local increase west of the Everglades is based on individuals self-reporting race as "Other," and Zhang, Lim, and Maiti (2014) results indicate this cluster is limited to the year 2000. Across the results we have a better description of the data pattern, allowing us to consider next steps such as confirming these suggested patterns by assessing the observed numbers of cases and children at risk to better describe the content and context of this collection of cases.
- Heaton's result identifying a shift in background rate in 2005 merits a closer look to see if reporting policies or practices changed at that time.

By summarizing findings across methods we obtain a more descriptive understanding of the case patterns in the data than we would have gotten with a simple "yes/no" answer to the question: Are there clusters of disease in the data? These more detailed descriptions provide important insight on data quality, shifting demographics of the at-risk population, and demographics of the observed disease cases. These are all important epidemiologic elements for understanding patterns of pediatric cancer in the state of Florida.

5. SUMMARY: HOW CLOSE ARE THE ANSWERED QUESTIONS TO THE MOTIVATING QUESTIONS?

The discussion above illustrates how each statistical method adds specificity to the general motivating question of interest. We also find that clearly stating these specifications add to our understanding of the results for each approach and patterns across approaches. This expanded understanding (moving past a simple yes/no answer) allows us to place results in clearer epidemiologic and policy contexts. While the results do not identify a "smoking gun" in the form of a shared environmental exposure in high-incidence areas, the results do provide epidemiologic insight into the local demographics of incident pediatric cancer cases, and suggest more detailed assessment of migration patterns in the Miami area. Policy-wise, the results point to responsibly responsive next steps of detailed description of the cases and the at-risk population in the detected areas to summarize local features in the data, particularly the race of cases west of the Everglades and demographic descriptors of any shifts in the at-risk population in the Miami area during the study period.

Again, I thank the authors for their insights and look forward to hearing more from each of them in future studies.

[Received February 2014. Revised January 2015.]

REFERENCES

Abrams, B., Anderson, H., Blackmore, C., Bove, F. J., Condon, S. K., Eheman, C. R., Fagliano, J., Haynes, L. B., Lewis, L. S., Major, J., McGeehin, M. A., Simms, E., Sircar, K., Soler, J., Stanbury, M., Watkins, S. M., and Wartenberg, D. (2013), “Investigating Suspected Cancer Clusters and Responding to Community Concerns: Guidelines From CDC and the Council of State and Territorial Epidemiologists,” *Morbidity and Mortality Weekly Report*, 62, 1–14. [3]

Alexander, F. E., and Boyle, P. (eds.) (1996), *Methods for Investigating Localized Clustering of Disease, IARC Scientific Publications No. 135*, Lyon, France: International Agency for Research on Cancer. [1]

Amin, R. W., Hendryx, M., Shull, M., and Bohnert, A. (2014), “A Cluster Analysis of Pediatric Cancer Incidence Rates in Florida: 2000–2010,” *Statistics in Public Policy*, 1, 69–77. [2,3,4,5,6,7]

Bender, A. P., Williams, A. N., Johnson, R. A., and Jagger, H. G. (1990), “Appropriate Public Health Responses to Clusters: The Art of Being Responsibly Responsive,” *American Journal of Epidemiology*, 132, S48–S52. [3]

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995), “Bayesian Computation and Stochastic Systems” (with discussion), *Statistical Science*, 10, 3–66. [6]

Besag, J., York, J. C., and Mollié, A. (1991), “Bayesian Image Restoration, With Two Applications in Spatial Statistics” (with discussion), *Annals of the Institute of Statistical Mathematics*, 43, 1–109. [6]

Best, N., Arnold, R. A., Thomas, A., Waller, L. A., and Conlon, E. M. (1999), “Bayesian Models for Spatially Correlated Disease and Exposure Data,” in *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford: Oxford University Press, pp. 131–156. [6]

Beyer, K. M. M., Schultz, A. F., and Ruston, G. A. (2008), “Using ZIP Codes as Geocodes in Cancer Research,” in *Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research, and Practice*, eds. G. A. Rushton, M. P. Armstrong, J. Gittler, B. R. Greene, C. E. Pavlik, M. M. West, and D. L. Zimmerman, New York: CRC Press, pp. 37–68. [4]

Coory, M. D., and Jordan, S. (2013), “Assessments of Chance Should be Removed From Protocols for Investigating Cancer Clusters,” *International Journal of Epidemiology*, 42, 440–447. [2]

Goodman, M., Naiman, N. S., Goodman, D., and LaKind, J. S. (2012), “Cancer Clusters in the USA: What do the Last Twenty Years of State and Federal Regulations Tell Us?” *Critical Reviews in Toxicology*, 42, 474–490. [3,7]

Grieco, E. M., and Cassidy, R. C. (2001), *Overview of Race and Hispanic Origin. Census 2000 Brief C2KBR/01-1*, Washington, DC: U.S. Census Bureau. [4]

Grubesic, T. H., and Matisziw, T. C. (2006), “On the Use of ZIP Codes and ZIP Code Tabulation Areas (ZCTAs) for the Spatial Analysis of Epidemiological Data,” *International Journal of Health Geographics*, 5, 58. [4]

Heaton, M. J. (2014), “Wombling Analysis of Childhood Tumor Rates in Florida,” *Statistics and Public Policy*, 1, 60–67. [2,4,5,6,7]

Lawson, A. B. (2006), *Statistical Methods in Spatial Epidemiology* (2nd ed.), Chichester: Wiley. [1]

Lawson, A. B., and Rotejanaprasert, C. (2014), “Childhood Brain Cancer in Florida: A Bayesian Clustering Approach,” *Statistics and Public Policy*, 1, 99–107. [2,3,4,5,7]

Neutra, R. R. (1990), “Counterpoint From a Cluster Buster,” *American Journal of Epidemiology*, 132, 1–8. [3]

Ren, C., Lim, S., Hylton, T., Huang, Y., Button, J., Wohler, B., Levin, G., and MacKinnon, J. (2012), *Florida Annual Cancer Report: 2008 Incidence and Mortality*, Tallahassee, FL: Florida Department of Health. [4]

Rothman, K. J. (1990), “A Sobering Start to the Cluster Busters’ Conference,” *American Journal of Epidemiology*, 132, S6–S13. [2]

Tango, T. (2010), *Statistical Methods for Disease Clustering*, New York: Springer. [1]

Thun, M. J., and Sinks, T. (2004), “Understanding Cancer Clusters,” *CA: A Cancer Journal for Clinicians*, 54, 273–280. [2,3,7]

Waller, L. A. (2009), “Detection of Clustering in Spatial Data,” in *The SAGE Handbook of Spatial Analysis*, eds. A. S. Fotheringham and P. A. Rogerson, London: SAGE. [1]

Waller, L. A., and Gotway, C. A. (2004), *Applied Spatial Statistics and Public Health Data*, Hoboken, NJ: Wiley. [1]

Wang, H., and Rodriguez, A. (2014), “Identifying Pediatric Cancer Clusters in Florida Using Loglinear Models and Generalized Lasso Penalties,” *Statistics and Public Policy*, 1, 96–86. [2,4,6,7]

Zhang, Z., Lim, C. Y., and Maiti, T. (2014), “Analyzing 2000–2010 Childhood Age-Adjusted Cancer Rates in Florida: A Spatial Clustering Approach,” *Statistics and Public Policy*, 1, 120–128. [2,4,5,6,7]