



NIH PUBLIC ACCESS

Author Manuscript

Health Place. Author manuscript; available in PMC 2011 May 1.

Published in final edited form as:

Health Place. 2010 May ; 16(3): 539–546. doi:10.1016/j.healthplace.2009.12.014.

THE SPATIAL STRUCTURE OF AUTISM IN CALIFORNIA, 1993–2001

Soumya Mazumdar [Postdoctoral Research Fellow],

Institute for Social and Economic Research and Policy, Columbia University, New York, New York, soumyamazumdar@yahoo.com +3472517975

Marissa King [Postdoctoral Research Fellow],

Institute for Social and Economic Research and Policy, Columbia University, New York, New York, mdk2101@columbia.edu +1 212 854 7530

Ka-Yuet Liu [Postdoctoral Research Fellow],

Institute for Social and Economic Research and Policy, Columbia University, New York, New York, kyl2111@columbia.edu: +1 212 854 7918

Noam Zerubavel [Research Assistant], and

Institute for Social and Economic Research and Policy, Columbia University, New York, New York, nz2104@columbia.edu @columbia.edu: +1 212 854 7918

Peter Bearman [Cole Professor of Social Science]

Institute for Social and Economic Research and Policy, Columbia University, New York, New York, Peter S. Bearman, psb17@columbia.edu*, +1 212 854 3094, Fax:+1 212-854-8925

Abstract

This article identifies significant high-risk clusters of autism based on residence at birth in California for children born from 1993 through 2001. These clusters are geographically stable. Children born in a primary cluster are at four times greater risk for autism than children living in other parts of the state. This is comparable to the difference between males and females and twice the risk estimated for maternal age over 40. In every year roughly 3% of the new caseload of autism in California arises from the primary cluster we identify – a small zone 20km by 50km. We identify a set of secondary clusters that support the existence of the primary clusters. The identification of robust spatial clusters indicates that autism does not arise from a global treatment and indicates that important drivers of increased autism prevalence are located at the local level.

Keywords

Autism; clusters; Spatial; GIS (Geographical Information Systems); California

© 2010 Published by Elsevier Ltd.

Correspondence to: Peter Bearman.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

Autism (DSM IV-299.0) (APA 1994) is a developmental disorder that impairs social interaction, communication and predisposes children to restrictive and repetitive behaviors. Between 1992 and 2006, the autism caseload in California increased by 598% (Newschaffer et al. 2007). Nor is California unique: where comparable data are available, comparable increases are observed (Newschaffer et al. 2007). These striking increases are associated with equally striking dissensus as to cause. While there is evidence of heritability, no marker for autism is associated with more than 10–15% of all cases, and prevalence has increased far too rapidly to be accounted for by fundamental changes in the human genome (Folstein and Rosen-Sheidley 2001). In the scientific community, diagnostic change and/or expansion, increased exposure to environmental toxins, demographic change, shifting prenatal and obstetric practices including the advent of new reproductive technologies, and complex social and environmental interactions with genetics are thought to play a role in the increased prevalence of autism (Ming et al. 2008, Palmer et al. 2006, Palmer and Wood 2009, Waldman et al. 2008, Windham et al. 2006, Ozand et al. 2003, Kolevzon et al. 2007, Glasson and Bower 2004). Although credible scientific studies provide no support, the belief that vaccinations cause autism is widely held in the lay community (Florida Institute of Technology 2008). This article identifies local spatial clusters of excess risk *at birth* for autism – thus pointing to the presence of local social and/or environmental factors as driving increased autism prevalence. By implication, it challenges arguments that propose causal agents that are distributed at random with respect to space.

Social and epidemiological research has proposed numerous individual-level risk factors as salient for autism (Glasson and Bower 2004, Kolevzon et al. 2007), yet the evidence for specific risk factors is often contradictory, in part arising from temporally and spatially heterogeneous study populations (King et al. 2009). Most critically, the majority of risk factors identified in the literature explain little of the overall variance. Two risk factors -- being male and having older parents -- have been established as definitive (Newschaffer et al. 2007, Schubert 2008). Evidence for the impact of other factors has been weak and inconclusive.

Historically, the discovery of spatial structures for diseases has been important for identifying mechanisms, and for falsification of competing hypotheses (Jacquez 2004, Buck 1975, Wagner 1980). If risk is spatially random or universal, then -- after adjusting for population heterogeneity with respect to known causes -- there will be no significant spatial variation in prevalence. Consequently, spatial clusters of increased autism should be absent (Fombonne 2003). On the other hand, the identification of local spatial clusters after adjustment for known causes provides powerful evidence in support of the existence of spatially non-random or non-universal etiological factors that could cause autism or autism diagnosis. Specifically, the observation of spatial clustering of autism cases at residence *at birth* indicates that a process that operates at a local scale is associated with amplified prevalence. In contrast, the observation of spatial clustering by residence *at diagnosis* could indicate selection into neighborhoods for services or the non-random diffusion of diagnostic regimes. Consequently, the spatial analysis of residence at birth is preferable to the spatial analysis of residence at diagnosis for identifying potential causes of autism.

Space, as a proxy for exposure, can also serve as a powerful exploratory tool for generating hypotheses. Once a reliable risk map is constructed, hypotheses can be generated about possible exposures. This is especially the case if clusters of significantly increased risk are observed. It is important that a cluster investigation be carried out with care and scientific rigor. All too commonly cluster investigations are driven by media reports, hearsay and anecdotal evidence. For example, the two “autism clusters” reported previously were identified on the basis of observations made by parents (Baron-Cohen et al. 1999, London and Etzel 2000). In the US

case, the perception of a cluster was the result of diagnostic misclassification and exhaustive case finding methods (Bertrand et al. 2001). In contrast, we use a statistically rigorous method – Kulldorff’s Spatial Scan Statistic (Kulldorff 1997), to demarcate clusters of high risk. Kulldorff’s SaTScan identifies a single statistically most likely “primary” cluster, and a list of secondary -- significant, but less likely -- clusters. We systematically search for clusters of sole autism (autism without other co-morbid conditions) for each year for the birth cohorts 1993 through 2001 in California. We map the clusters and their spatial stability over time.

Methods

Data

We use case and control data obtained by exact and probabilistic matching of all persons with autism served by the California Department of Developmental Services (DDS) during the period from 1993 to 2005 (11,683 cases) to the California Birth Master Files (BMF) for the birth cohorts 1993–2001 (4,176,783 births). Matches were made based on first, middle, and last name, sex, race, date of birth, and maternal zip code at birth. We focus on the birth cohorts of 1993 to 2001 in order to eliminate temporal ascertainment biases. Cases of autism (ICD-10 F84.0) not co-morbid with mental retardation or “sole autism” were extracted from this dataset, aggregated by year and geo-coded to Census-Bureau Zip Code Tabulation Areas (ZCTAs) for the years 1993 to 1996 and to census block groups for the years 1997 to 2001. The BMF has zip codes at birth through 1996 and individual addresses from 1997 forward. The addition of addresses at birth to the Birth Master Files from 1997 onwards allowed for greater geographic precision in the later birth cohorts. Approximately 93% of the records were geo-coded successfully. The BMF comprise a control population which is used to calculate the expected number of cases.

Method

Kulldorff’s Spatial Scan Statistic (Kulldorff 1997), is implemented by the SaTScan Software (Kulldorff 2006). Kulldorff’s Spatial Scan Statistic (Kulldorff 1997) reduces the problem of cluster detection to a problem of maximum likelihood estimation over geographical space. Through a single hypothesis test over space, the method identifies a region, where the distribution of cases relative to controls (or the expected number of cases) is most likely to be consistent with a significant excess of risk. By carrying out just one hypothesis test over the entire geographical space, this method solves the problem of multiple hypothesis testing that has plagued the cluster detection literature. SaTScan identifies candidate clusters, which are circles of increasing radii, bound by a maximum threshold radius, centered on pre specified locations such as ZCTA centroids. Over many candidate clusters SaTScan maximizes the likelihood ratio, given by

$$LLR = O * \ln \left(\frac{O}{E} \right) + O * \ln \left(\frac{n - O}{n - E} \right)$$

where, LLR represents the logarithm of the likelihood ratio, O is observed cases, E is expected cases, and n is the total number of cases in the entire region (California). The likelihood formula assumes that autism cases are distributed as a Poisson random variable and the likelihood ratio is compared to simulated likelihood ratios generated from Monte Carlo randomizations of the data to assess statistical significance. The area that has the highest likelihood value (or the lowest p value) is the primary cluster.

In the absence of any known risk factors, the expected number of cases is calculated by multiplying the California case rate with the control population in the circle. If there are any

known risk factors that are not spatially random, then an indirect standardization (Armstrong 1969, Julious et al. 2001) can be done. We standardize (Julious et al. 2001, Armstrong 1969) for parental age, a spatially structured risk factor. We thus calculate what the expected number of cases in a neighborhood (block group or ZCTA) would be if the rates in a given age group in the neighborhood were the same as observed for that age group in all of California. For each birth, we use a single categorical parental age variable for the adjustment process. The categorical parental age variable is derived from a continuous parental age variable, by grouping the continuous variable into 35 and older and 34 and younger age categories. We analyzed the data for each year from 1993 to 2001 individually searching for non-overlapping clusters of high and low risk of sole autism in California.

The choice of an appropriate maximum threshold radius for the cluster or maximum cluster size is important. For this analysis, the maximum allowable cluster size was limited 1% of the California population, or roughly 5000 births. A 1% upper bound on cluster size provides locational certainty without compromising statistical power. Smaller clusters have more statistical and spatial uncertainty, but they allow us to zoom in on specific neighborhoods (Silverman 1978, Silverman 1986). Larger clusters are more robust, but provide less geographic information. One approach around this problem is to map nested clusters at various spatial scales (Boscoe et al. 2003, Boscoe 2008, Chen et al. 2008). The existence of nested clusters at local scales suggests risks that decay from the center of the nest outwards. A smaller SaTScan cluster that nests in multiple larger clusters is also statistically reliable, and is called a “core cluster” by some researchers (Boscoe 2008, Chen et al. 2008). We mapped age adjusted spatial clusters at 1%, 5% and 10% maximum population thresholds. For all years (1993–2001), the 1% clusters are nested within the 10% clusters. For a majority of years the 1% clusters nest within or have a high degree of overlap with the 5% clusters. The risks in the 1% clusters are always higher than observed in the 5% and 10% clusters. Figure 1 displays the nested clusters for three representative years and the risks in the nested clusters at all years. From these analyses we observe that the 1% scan identifies neighborhoods with the highest elevation of risk. Thus, a 1% threshold is most likely to identify local and reliable (Boscoe 2008, Chen et al. 2008) clusters. Since we execute separate scans for each year, and produce maps of temporally stable clusters, we control for geographic uncertainty even though we use a small cluster population threshold.

To evaluate the public health impact of a cluster, it is possible to calculate the population attributable fraction (PAF) -- defined as the proportional reduction in risk gained by elimination of the exposures (Rockhill et al. 1998, Yiannakoulias 2009). We use the following formula:

$$\sum_{i=1}^k (O_i/n) * ((RR_i - 1)/RR_i)$$

Where O_i is the observed number of cases in region i , n is the total number of cases in California over the observation window, RR_i is the relative risk in the i 'th cluster, with $i=1$ the primary cluster, and $i=2, i=3, i=n$ for the secondary clusters, up to k secondary clusters.

We check the robustness of our methods, assess whether the aggregation of data biases our analyses and ensure that the reference distributions we use are valid. To test for aggregation biases from the Modifiable Area Unit Problem (Openshaw and Taylor 1979) we use a one in five sample from individual level data for 1997 (the middle of our study period) and a Bernoulli (Kulldorff 1997) Spatial Scan Model to look for clusters. We leave the other Spatial Scan parameters the same as the Poisson model. The results from the two models are consistent. The most likely cluster of the Bernoulli model shares 88% of its area with most likely cluster found using the Poisson model. The default number of Monte Carlo randomizations used by SaTScan

to evaluate statistical significance is 999. We use this default in our analyses, but we also test the validity of these reference distributions. For three representative years, we simulated 9999 likelihood ratios to generate the reference distributions. The most likely clusters that were significant with 999 simulations remained so with 9999 simulations with the p values converging towards zero with a greater number of simulations. We therefore conclude that 999 simulations are sufficient to generate the reference distributions. Our conclusion is supported by extensive power calculations undertaken by the creators of the Spatial Scan Statistic, where 999 simulations were found sufficient to generate robust reference distributions (Kulldorff et al. 2003).

Results

A significant primary cluster is found for every birth cohort in every year. All clusters of sole autism are located within a 50 km by 20 km area of Northern Los Angeles centered on West Hollywood. Within this confined area, there is spatial overlap between the clusters that appear for each year (Figure 2). Children born in these neighborhoods are at approximately four times greater risk of autism than those born in any other place in California (Column 1, Table 1). Relative to known risk factors, this is comparable to the difference between males and females and is twice the risk estimated for maternal age over 40 (Newschaffer et al. 2007). It is possible that birth clusters -- clusters defined by residence of the child at birth -- arise from compositional dynamics. Parents at risk for a child with autism may select communities to reside in, and then have children. Increased parental age and being male are known risk factors for autism and could drive neighborhood selection. That is, parents with preferences for male children could select specific kinds of neighborhoods to give birth to their child. There is no evidence that neighborhoods vary by proportion of male births, and therefore there is no evidence that neighborhoods are selective on parents who have gender preferences for their children. Likewise, older parents may have residential preferences different than those of younger parents. This is the case as neighborhoods vary by age of parents. Nonetheless, adjusting for parental age does not affect the location of the clusters as can be seen in Figure 3, and column 2 of Table 1.

The public health impact of the clusters is measured by their PAF. In our analysis, the PAF ranges between a low of 2.4% in 1996 to a high of 4.4% in 1993 (Column 5, Table 1). The PAF is 2.83% for all years. Thus if the primary clusters we identify were absent, the California sole autism caseload would decrease by close to 3%. Given that the primary clusters are bounded within a 20km × 50km area, and account for less than 1% of all births in California at any time over the observation window, this is a striking contribution.

We identify a set of statistically less likely set of secondary high-risk clusters (Figure 4). The secondary clusters are located in close proximity to the primary clusters. When all significant clusters (primary and secondary) are grouped together, the PAF ranges from 4.4% in 1993 to 15.1% in 1998 (Column 6, Table 1). The PAF for primary and secondary clusters combined is 11.7% for all years.

The etiological conditions in the primary clusters are spatially correlated with those of its neighboring regions, and therefore, secondary clusters are more likely to present themselves at a close proximity to the primary clusters than elsewhere. This is explained as a “bleed off” effect of statistical power from the primary clusters (Kulldorff 1997). In contrast to the primary cluster, some of the smaller secondary clusters are temporally unstable, and while increased risk is observed, they may be a product of chance.

Secondary clusters are less likely than primary clusters. They are more numerous than the primary clusters and therefore give the appearance of covering a large geographical area.

Nevertheless, their proximity to the primary clusters, and the absence of any secondary clusters in other parts of California underscore the statistical likelihood of the primary clusters.

An interesting secondary cluster, for the 1994 birth cohort, is a cluster of 12 cases (1.4 expected) centered on the neighborhood of Northridge (Relative Risk=8.7). In 1994, Northridge was the epicenter of the North Ridge Earthquake, and while tenuous, there is research that points to a relationship between increased maternal stress from large scale natural disasters during pregnancy and autism risk (Kinney et al. 2007). The other secondary clusters of interest are the most likely clusters of low risk. The most likely secondary clusters of low risk are located in sparsely populated areas near San Diego. This is an area geographically and compositionally distinct from the clusters of high risk. The age adjusted most likely secondary clusters of low risk are located in the sparsely populated desert areas in the vicinity of San Diego for the 1996 to 2001 birth cohorts. In a typical year, a child born in these neighborhoods have a 200% to 300% less risk of being born with sole autism than in the rest of California.

The neighborhoods that are part of the primary clusters can be compared to neighborhoods that are not part of primary clusters. We define the boundary of the primary clusters as the outer boundary of the dark blue region in Figure 3. An appropriate comparison region is a 3 km geographically contiguous area around the primary cluster. This is represented in Figure 5.

The comparison region (hereafter, buffer zone) has the same area and urbanicity/population density (~3000 people/sq km) as the primary cluster. The median property value in the cluster ranges from \$300,000 to \$499,999, while in the buffer zone it ranges from \$150,000 to \$174,999. In addition, 62% of the cluster population -- compared to 41% of the buffer zone population -- is white. Thus the primary cluster neighborhoods are socioeconomically different from the buffer zones immediately around them. Primary clusters are associated with more autism advocacy organizations (in 1993 and beyond) than expected by chance distribution. Primary clusters are not related at any time to DDS Regional Center catchment areas.

Discussion

We identify temporally robust, statistically significant clusters. However, in our approach, the spatial filter utilized by SaTScan is limited to circular geometries. Since, clusters can have various shapes; this geometrical constraint can decrease statistical power, and increase true negatives (Duczmal and Assuncao 2004). This, and the fact that we map the temporal stability of the clusters may render our results less sensitive but more conservative and specific (Duczmal et al. 2006, Kulldorff et al. 2006, Tango and Takahashi 2005). Since cluster alarms are a serious public health concern (Fombonne 2003), it is required of us that we be conservative. In addition, the sensitivity of a cluster search is dependent not only on the method used, but also the granularity of the data (Ozonoff et al. 2007). From 1997 to 2001, our data are at the block group level, which provide sensitivity to these analyses without increasing false positives.

Since the clusters are birth clusters, they are not subject to significant temporal diagnostic ascertainment bias. Area effects could be generated by (a) compositional effects, (b) peculiarities of the physical environment, such as the presence of toxins, or (c) the social environment, such as community cohesion or disorganization, or the interaction of some or many of these elements. It is possible that families with an older sibling diagnosed with autism move to areas where services are provided and subsequently give birth to a child at higher risk for autism. If this were the case, selection of neighborhoods could play a role in the generation of the primary cluster. An analysis of moves into the primary cluster we identify does not support this hypothesis. While families with children with autism move more than comparable families, moves tend to be short and within cluster boundaries. What cannot account for these

local clusters are risks that are globally distributed without the identification of specific local environmental interactions as a central causal mechanism.

Conclusion

Cluster reports, even when carried out with scientific rigor should be evaluated with caution (Kingsley et al. 2007). The discovery of a pronounced spatial structure for autism suggests that local environmental or social dynamics play a role in autism risk, but do not point precisely to the causal process involved. Thus, cluster detection should be used to disprove rather than confirm causality (Jacquez 2004). Here we identify clusters of significant high risk at a fine resolution that are geographically stable over long periods of time. This indicates that our search for the etiology of increased autism prevalence will be facilitated by focus on very local processes.

It is possible that institutional diagnostic dynamics induce the clusters we observe. Specifically, diagnosticians in the regional centers of the California Department of Developmental Services (DDS) who serve the catchment areas for our clusters may be more likely to diagnose individuals with autism in order to provide services than their counterparts servicing other catchment areas. If this were the case, we would expect to observe that individuals diagnosed with autism arising from the LA centers would – on average – be diagnosed as higher functioning than those individuals arising from other DDS offices. This is not the case. With respect to a global functioning score that encompasses repetitive behavior, communication and social skills the two populations are equivalent. As institutional diagnostic dynamics are not on the surface associated with increased autism prevalence within the LA cluster area, social or environmental dynamics provide the most likely explanation.

Other institutional processes may be influential. The activity of advocacy organizations (lagged one year for causal order) clearly plays some role in diffusing knowledge about autism, the interpretation of behavioral symptoms, and may provide parents with resources and information that assists them in their efforts to secure a diagnosis for their child. Likewise, parents quite likely diffuse information through informal networks. If these dynamics are operating, even very small environmental risks could yield the amplification of autism risk we observe in the primary cluster areas. This requires a multivariate approach to study the interplay of various, social, environmental and biological risk factors, and how they shape the autism epidemic. Thus the next critical step is to incorporate multiple parameters into a simulated complex world (Epstein 2009), where the emergence of the autism epidemic can be interactively modeled and observed.

Acknowledgments

This research is supported by the NIH Director's Pioneer Award program, part of the NIH Roadmap for Medical Research, through grant number 1 DP1 OD003635-01.

REFERENCES

- Armstrong RW. Standardized Class Intervals and Rate Computation in Statistical Maps of Mortality. *Annals of the Association of American Geographers* 1969;59:382–390.
- Association, A. P. *Diagnostic and statistical manual of mental disorders, DSM IV*. Washington: American Psychiatric Publishing; 1994.
- Baron-Cohen S, Saunders K, Chakrabarti S. Does autism cluster geographically? A research note. *Autism* 1999;3:39–43.
- Bertrand, J.; Mars, A.; Boyle, C.; Bove, F.; Yeargin-Allsopp, M.; Decoufle, P. Prevalence of Autism in a United States Population. Vol. 108. *The Brick Township, New Jersey: Investigation. Pediatrics*; 2001. p. 1155-1161.

- Boscoe FP. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality (Comment). *International Journal of Health Geographics* 2008;7:1. [PubMed: 18190678]
- Boscoe FP, McLaughlin C, Schymura MJ, Kielb CL. Visualization of the spatial scan statistic using nested circles. *Health and Place* 2003;9:273–277. [PubMed: 12810334]
- Buck C. Popper's philosophy for epidemiologists. *International Journal of Epidemiology* 1975;4:159–168. [PubMed: 171230]
- Chen J, Roth RE, Naito AT, Lengerich EJ, MacEachren AM. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *International Journal of Health Geographics* 2008;7:1–18. [PubMed: 18190678]
- Duczmal L, Assuncao R. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics and Data Analysis* 2004;45:269–286.
- Duczmal L, Kulldorff M, Huang L. Evaluation of spatial scan statistics for irregularly shaped clusters. *Journal of Computational and Graphical Statistics* 2006;15:428–442.
- Epstein J. Modelling to contain pandemics. *Nature* 2009;460:687. [PubMed: 19661897]
- Florida Institute of Technology. American's Knowledge and Understanding of Autism, 2008. Melbourne, Florida: 2008.
- Folstein SE, Rosen-Sheidley B. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nature Reviews Genetics* 2001;2:943–955.
- Fombonne E. Epidemiological surveys of autism and other pervasive developmental disorders: An update. *Journal of Autism and Developmental Disorders* 2003;33:365–382. [PubMed: 12959416]
- Glasson EJ, Bower C. Perinatal factors and the development of autism. *Archives of Pediatric and Adolescent Medicine* 2004;61:618–627.
- Jacquez GM. Current practices in the spatial analysis of cancer: flies in the ointment. *International Journal of Health Geographics* 2004;3:1–10. [PubMed: 14748927]
- Julious SA, Nicholl J, George S. Why do we continue to use standardized mortality ratios for small area comparisons? *Journal of the Public Health Medicine* 2001;23:40–46.
- King M, Fountain C, Dakhllallah D, Bearman PS. Estimating autism risk in a time of increasing reproductive age. *American Journal of Public Health* 2009;99:1673–1679. [PubMed: 19608957]
- Kingsley BS, Schmeichel KL, Rubin CH. An update on cancer cluster activities at the Centers for Disease Control and Prevention. *Environmental Health Perspectives* 2007;115:165–171. [PubMed: 17366838]
- Kinney DK, Miller AM, Crowley DJ, Huang E, Gerber E. Autism prevalence following prenatal exposure to hurricanes and tropical storms in Louisiana. *Journal of Autism and Developmental Disorders* 2007;38:481–488. [PubMed: 17619130]
- Kolevzon A, Gross R, Reichenberg A. Prenatal and perinatal risk factors for autism: A review and integration of findings. *Archives of Pediatric and Adolescent Medicine* 2007;161:326–333.
- Kulldorff M. Spatial Scan Statistic. *Communications in Statistics: Theory and Methods* 1997;26:1481–1496.
- Kulldorff, M. SaTScan™ v7.0: Software for the spatial and space-time scan statistics. Bethesda: Information Management Services, Inc; 2006.
- Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic scan statistic. *Statistics in Medicine* 2006;25:3929–3943. [PubMed: 16435334]
- Kulldorff M, Tango T, Park PJ. Power comparisons for disease clustering tests. *Computational Statistics and Data Analysis* 2003;42:665–684.
- London E, Etzel RA. The environment as an etiologic factor in autism: A new direction for research. *Environment Health Perspectives* 2000;108:401–404.
- Ming X, Brimacombe M, Malek JH, Jani N, Wagner GC. Autism spectrum disorders and identified toxic landfills: co-occurrence across states. *Environmental Health Insights* 2008;2:55–59.
- Newschaffer CJ, Croen LA, Daniels J, Giarelli E, Grether JK, Levy SE, Mandell DS, Miller LA, Pinto-Martin J, Reaven J, Reynolds AM, Rice CE, Schendel C, Windham GC. The epidemiology of autism spectrum disorders. *Annual Review of Public Health* 2007;28:235–258.

- Openshaw, S.; Taylor, PJ. A million or so correlation coefficients: Three experiments in the modifiable area unit problem. In: Wrigley, N., editor. *Statistical Applications in the Spatial Sciences*. London: Pion; 1979.
- Ozand PT, Ali-Odaib A, Merza H, Al Harbi S. Autism: a review. *Journal of Pediatric Neurology* 2003;1:55–67.
- Ozonoff A, Jeffery C, Manjourides J, Forsberg White L, Pagano M. Effect of spatial resolution on cluster detection: a simulation study. *International Journal of Health Geographics* 2007;6:1–7. [PubMed: 17214903]
- Palmer RF, Blanchard S, Steina Z, Mandell D, Miller C. Environmental mercury release, special education rates, and autism disorder: an ecological study of Texas. *Health and Place* 2006;12:203–209. [PubMed: 16338635]
- Palmer RFBS, Wood R. Proximity to point sources of environmental mercury release as a predictor of autism prevalence. *Health and Place* 2009;15:18–24. [PubMed: 18353703]
- Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *American Journal of Public Health* 1998;88:15–19. [PubMed: 9584027]
- Schubert C. Male biological clock possibly linked to autism, other disorders. *Nature Medicine* 2008;14:1170.
- Silverman BW. Choosing the window width when estimating a density. *Biometrika* 1978;65:1–11.
- Silverman, BW. *Density estimation for statistics and data analysis*. Boca-Raton: Chapman and Hall/CRC; 1986.
- Tango T, Takahashi K. A flexibly shaped scan statistic for detecting clusters. *International Journal of Health Geographics* 2005;4:1–15. [PubMed: 15649329]
- Wagner GC. The detection of carcinogenic effects in man. *Journal of Research in Clinical Oncology* 1980;97:215–221.
- Waldman M, Nicholson S, Adilov N, Williams J. Autism prevalence and precipitation rates in California, Oregon, and Washington counties. *Archives of Pediatric and Adolescent Medicine* 2008;162:1026–1034.
- Windham GC, Zhang L, Gunier RB, Croen LA, Grether JK. Autism spectrum disorders in relation to distribution of hazardous air pollutants in the San Francisco bay area. *Environmental Health Perspectives* 2006;114:1438–1444. [PubMed: 16966102]
- Yiannakoulis N. Using population attributable risk to understand geographic disease clusters. *Health & Place*. 2009 In Press.

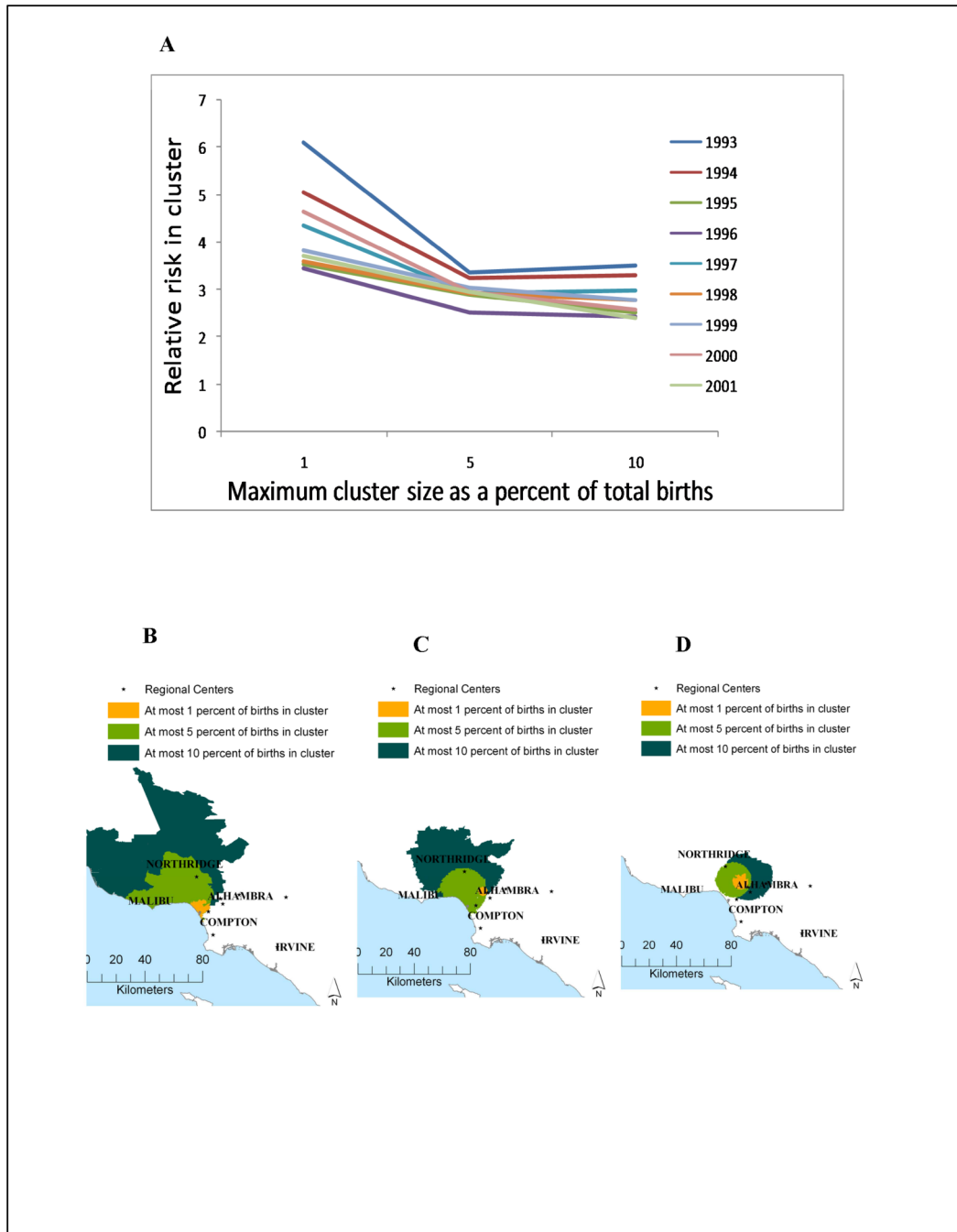


Figure 1. Choosing an appropriate cluster size through nested visualization of scan results
 (A) Risk decays as cluster size becomes larger. The effect is generally consistent through the years. Nested clusters for the years (B) 1994 (C) 1997 and (D) 2000. The larger clusters nest the smaller ones.

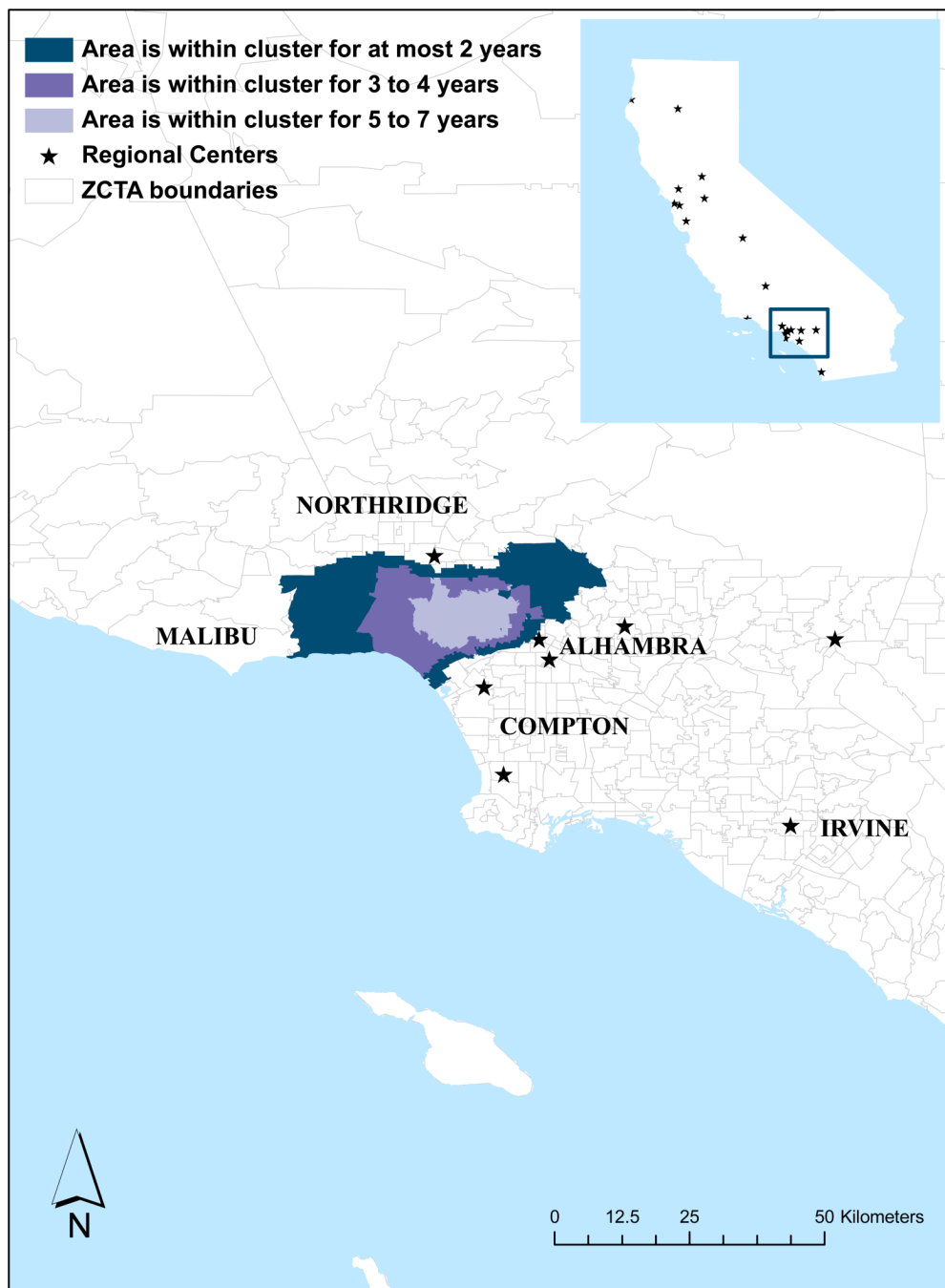


Figure 2. Most likely or primary high risk clusters of sole autism

This figure displays the number of years a given block group belongs to a cluster. Concentric rings show a central core, with block groups at the center belonging to a cluster more often than block groups on the periphery.

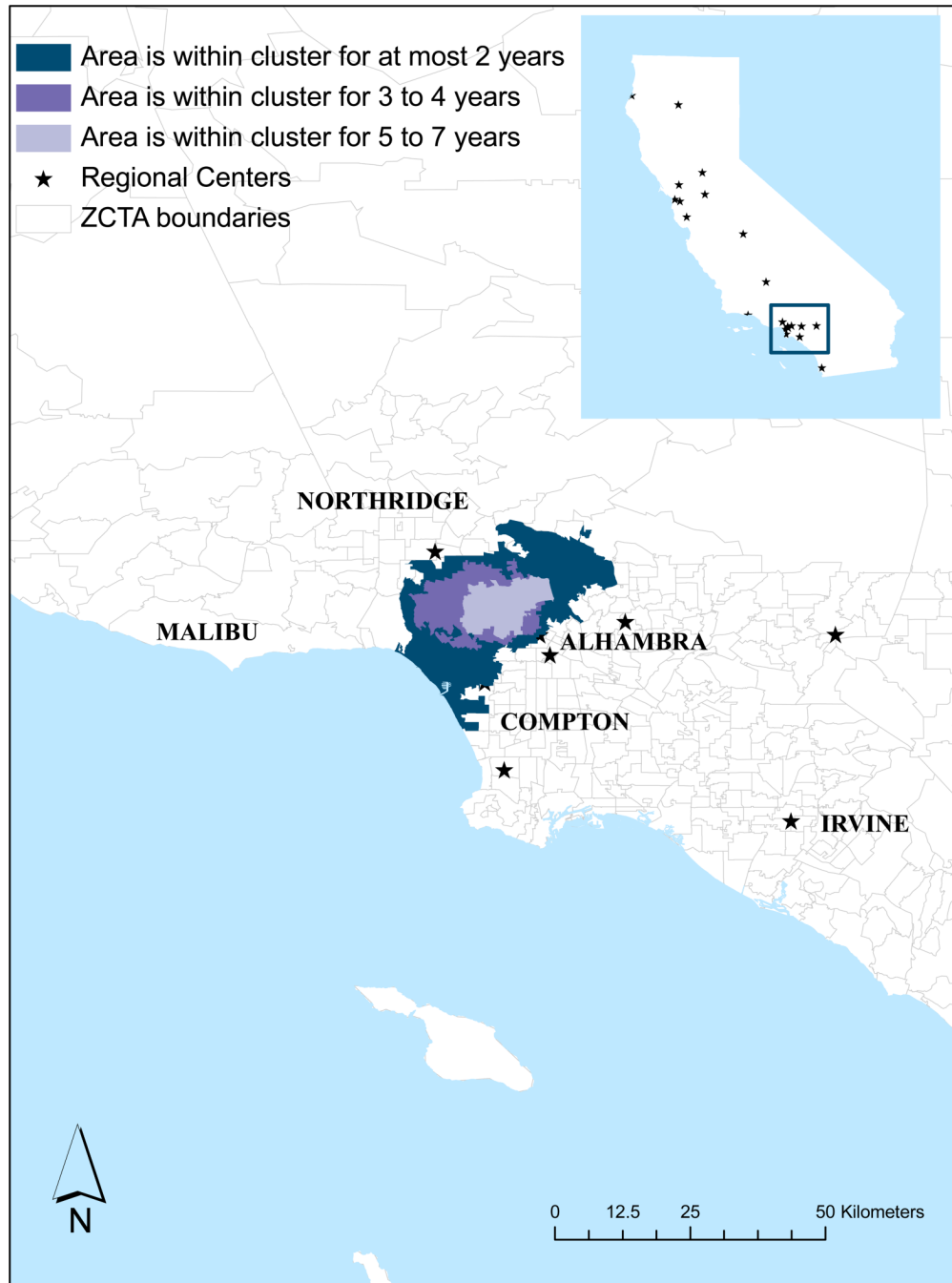


Figure 3. Most likely or primary high risk clusters of sole autism after adjusting for age

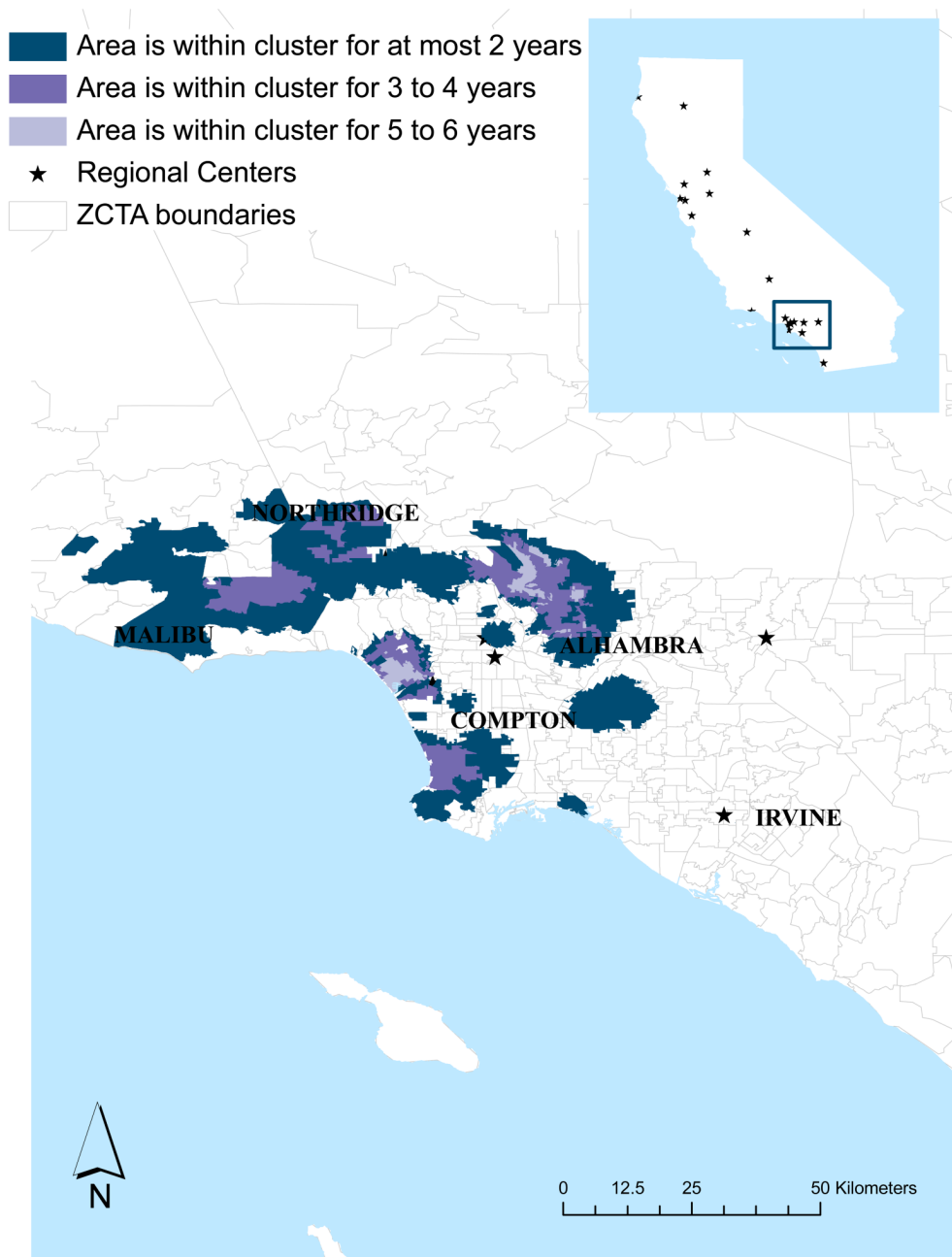


Figure 4. Age adjusted secondary clusters of high risk
 There are 38 age adjusted secondary clusters in the time period 1993–2001 with a median risk of 3.2.



Figure 5. Choosing an appropriate comparison neighborhood
 A 3 km buffer around the “cluster” region forms an appropriate comparison neighborhood

Table 1

Attributes of clusters before and after adjusting for age.

Year	Risk at most likely cluster before adjusting for age	Risk at most likely cluster after adjusting for age	Observed number of cases in age adjusted clusters	Expected number of cases after adjustment	Population attributable risk proportion for age adjusted primary clusters	Population attributable risk proportion for both primary and secondary age adjusted clusters
1993	6.2	6.1	**	**	4.38	4.38
1994	5.3	5.0	**	**	3.06	10.29
1995	4.4	3.5	**	**	2.44	8.45
1996	3.8	3.5	34	10	2.39	7.78
1997	4.7	4.4	56	13	2.82	14.26
1998	4.1	3.6	51	15	2.51	15.10
1999	4.2	3.8	53	14	2.61	10.34
2000	4.9	4.6	78	17	3.47	13.45
2001	3.8	3.7	59	16	2.73	10.92

Note that after adjusting for age, the cluster may not be located in the same spatial location as before adjustment. All clusters are significant with $p < 0.001$.

** Data are suppressed to conform to Institutional Review Board requirements since observed number of cases are less than 30.