NIH Public Access
Author Manuscript
Autism Res. Author manuscript; available in PMC 2014 June 25.

Published in final edited form as:
Autism Res. 2010 February ; 3(1): 19–29. doi:10.1002/aur.110.

Geographic Distribution of Autism in California: A Retrospective Birth Cohort

Karla C Van Meter[1,2], **Lasse E Christiansen**[2,3], **Lora D Delwiche**[1], **Rahman Azari**[4], **Tim E Carpenter**[2,5], and **Irva Hertz-Picciotto**[1,6]

[1]Department of Public Health Sciences, School of Medicine, and M.I.N.D. (Medical Investigations of Neurodevelopmental Disorders) Institute, University of California, Davis USA

[2]Center for Animal Disease Modeling and Surveillance, School of Veterinary Medicine (CADMS), University of California, Davis USA

[3]Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby DK

[4]Department of Statistics, University of California, Davis USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis USA

[6]Center for Children's Environmental Health, School of Veterinary Medicine, University of California, Davis USA

Abstract

Prenatal environmental exposures are among the risk factors being explored for associations with autism. We applied a new procedure combining multiple scan cluster detection tests to identify geographically defined areas of increased autism incidence. This procedure can serve as a first hypothesis-generating step aimed at localized environmental exposures, but would not be useful for assessing widely distributed exposures, such as household products, nor for exposures from non-point sources, such as traffic.

Geocoded mothers' residences on 2,453,717 California birth records, 1996–2000, were analyzed including 9,900 autism cases recorded in the California Department of Developmental Services (DDS) database through February 2006 which were matched to their corresponding birth records. We analyzed each of the 21 DDS Regional Center (RC) catchment areas separately because of wide variation in diagnostic practices. Ten clusters of increased autism risk were identified in eight RC regions, and one potential cluster in each of two other RC regions.

After determination of clusters, multiple mixed Poisson regression models were fit to assess differences in known demographic autism risk factors between births within and outside areas of elevated autism incidence, independent of case status.

Adjusted for other covariates, the majority of areas of autism clustering were characterized by high parental education, e.g., relative risks >4 for collegegraduate versus non-high school graduate parents. This geographic association possibly occurs because RCs do not actively conduct case

Corresponding author: Karla Van Meter, vanmeter@pobox.com, Mobile: 1-707-483-0420, Fax: 1- 530-752-1618, UC Davis, VM: CADMS, One Shields Avenue * 1044 Haring Hall, Davis CA 95616 USA.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

finding and parents with lower education are, for various reasons, less likely to successfully seek services.

Keywords

autism; cluster; environmental; epidemiology; scan tests; spatial; geographic; sociodemographic

Introduction

Analysis of spatial patterns of disease is often used as a first step toward identifying environmental factors that cluster geographically and hence cause elevated incidence rates. While many environmental exposures might be widespread (e.g., medications, chemicals in widely used household products or traffic pollution), pollution point sources such as factories, local water systems, and waste sites do result in more geographically restricted dispersion of contaminants. Conversely, if cases of a disease tend to show a clustered pattern, pollutants from local sources might be of interest.

Three previous studies have examined the relationship of specific geographically defined environmental exposures with autism. Two case-control studies with individual-level data on diagnoses and confounders evaluated autism status in relation to chemical exposure estimates modeled using government databases. Windham et al. (2006) examined 1994 births in the San Francisco Bay Area to assess associations with the 1996 modeled hazardous air pollutant concentrations, defined by census tract. Roberts et al. (2007) used records of pesticide applications by weight from the California Department of Pesticide Regulation to evaluate autism risk in relation to estimated prenatal exposures. Palmer's (2005) ecological study in Texas examined the association of mercury releases with rates of autism by school district.

Given the large number of candidate environmental exposures and the low (though rising) prevalence of autism, rather than focus on any specific set of chemicals, we took a different approach. Specifically, we undertook this study to search for geographic areas with significantly increased incidence of autism among births in the area, i.e., clusters.

Cluster detection tests (CDTs) have been developed to identify clustering of an event in time or space that is not likely random. Spatial scan tests are among the most powerful CDTs for geographically locating statistically significant clusters of events. They have been used to assess whether reported 'clusters' of disease such as breast cancer are likely to be due to random variation (Kulldorff et al. 1997), and for more basic exploration of disease geography (Christiansen et al. 2006).

CDTs can be used either as a preliminary method to explore for spatial clusters without hypothesizing a specific risk factor or to assess the statistical significance of locally increased event numbers that are a public concern. Identified clusters where the increase cannot be explained by geographic clustering of known risk factors may be locations where a further focused environmental analysis would be worthwhile. At that point, exposures of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

highest concern will be locally occurring ones, especially if they are elsewhere uncommon or present at lower levels.

Before embarking on such focused investigation, it may be advisable to evaluate whether spatial clustering of elevated autism incidence is confounded by clustering of known demographic risk factors. The known demographic risk factors for autism that may also cluster geographically are advanced maternal and paternal ages and parental education (Bhasin and Schendel (2007), Croen et al (2002), Croen et al. (2007), Glasson et al. (2004), Juul-Dam et al. (2001), Lauritsen et al. (2005), Reichenberg et al. (2006)). Many authors suggest that the parental education association may be one of case-ascertainment. Bhasin and Schendel also indicated an ascertainment association with race. Elevated associations of these demographic variables with the cluster birth populations would reduce the likelihood of independent point-source environmental exposures being responsible for the autism clustering found.

While there are tests that can evaluate risk factors by stratifying the population on potential confounding variables, the low incidence of rare diseases means that each cluster may have few affected individuals, severely limiting the number of strata that can be used. Autism is sufficiently rare that, for example, two siblings with autism could constitute a highly significant spatial cluster in a certain CDT. Stratification on multiple demographic risk factors simultaneously would quickly create dozens of strata. With a median of 504 affected individuals per RC catchment area, stratification was not feasible.

An alternative to stratification builds on the understanding that a risk factor can only be a confounder if it were associated with outcome (autism) *and* exposure (spatial proximity). We therefore fit multivariate Poisson regression models where the outcome was birth in an area of autism clustering (identified using the multiple CDT tests) and the predictors were the demographic factors already known to be associated with autism. The median cluster size was 10,166 births.

We previously developed a procedure to improve the specificity and sensitivity of cluster detection of rare disorders using unstratified multiple CDTs (Van Meter et al., 2008). The current study is an application of that procedure to the 1996–2000 birth cohort of California. We conducted a search for clusters of autism, and followed with a statistical analysis of the association between those clusters and a set of known demographic risk factors for autism.

The source of our study's autism cases was the California Department of Developmental Services (DDS), which funds statewide services for people with developmental disabilities. Clients of the DDS, both with autism and milder Autism Spectrum Disorder (ASD), must have "substantial disabilities" with "significant functional limitations." The DDS administers a statewide system of 21 independent Regional Centers (RCs), where eligibility for services is determined.

As independent organizations, the RCs were not required to use the same diagnostic practices during the study period. For example, they varied in their use of independent clinical psychologist providers versus in-house clinical psychologists to complete the diagnostic evaluation as well as the use of school, psychiatric and pediatric reports in the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

process. Because of this variation in diagnostic practices, direct comparisons across RC regions would be invalid. We therefore analyzed each RC region separately to identify clusters.

Materials and methods

This study was approved by the Institutional Review Boards for the Protection of Human Subjects of the University of California, Davis and the State of California.

Cohort information

To link incident cases of autism to the cohort of all births, we matched records from the California state birth registry to the administrative data system of the California DDS. Records of all live births in California occurring in 1996–2000 (n= 2,634,527) were obtained from the California Department of Public Health's Office of Health Information and Research; we augmented the electronic Birth Statistical Master Files (*Center for Health Statistics. Confidential Birth 980-Byte File*, 1996–2000), with 1996–1997 Automated Vital Statistics System files. Variables from those records included both parents' ages, years of education, race, and ethnicity, birth type, and mother's address. Using ArcGIS 9 (ESRI, Redlands CA), the mother's address at time of delivery was geocoded successfully for 93 percent of records. We take this address as a point approximation for late gestational or early neonatal exposures.

After cleaning, we constructed analysis variables. Education was categorized based on highest level of education completed by one or the other parent, with four similarly sized groups: less than high school-graduate (less than 12 years), high school graduate (12 years), some college (13 to 15 years) and college graduate or above (greater than or equal to16 years.). The cohort included 2.6 percent multiple births, almost all twins.

The four variables for self-described race and ethnicity of the two parents were combined into a single three-level child race/ethnicity variable: Hispanic (at least one parent was ethnically Hispanic and both were either white or unknown race); white, non-Hispanic (at least one parent was white, the second white or unknown race and neither was ethnically Hispanic); other (at least one parent had a nonwhite race). Approximate distribution was: one-half Hispanic, one-quarter white non-Hispanic, and one quarter 'other;' Children with both parents of unknown race and not Hispanic were excluded from further analysis (n=9,410).

Parental ages, in years, were kept as continuous variables as preliminary analysis of both mother's and father's ages showed fairly smooth increasing relationships to autism cumulative incidence throughout the age ranges. We excluded births if the mother's age was outside the range of 10 through 55 years (n=25), if father's age was outside 10 through 74 years (n=76), or age was less than 4 years above the reported years of education (n=337 mothers, n=101 fathers). Less than one percent of autism cases had mothers or fathers outside the 16 to 45 year age range or the 16 to 55 year age range, respectively.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Validity of birth certificate information varies, with demographic data of high reliability as compared with, e.g., medical conditions (DiGiuseppe, Aron, Ranbom, Harper, & Rosenthal, 2002; Roohan et al., 2003).

Case information

Eligibility for DDS services is determined at each of the 21 RCs based on professional diagnostic evaluation. (Milder ASD cases are eligible for DDS services only when also substantially developmentally disabled.) Clients seek services on the advice of their child's pediatrician, teacher, or other sources. Croen et al. (2002) estimated that 75 to 80 percent of all California children with autism are included in the DDS records.

The DDS Client Development Evaluation Report (CDER) for each eligible child includes all diagnoses used to qualify for services, using United States standard morbidity ICD-9-CM codes (*International Classification of Diseases Ninth Revision, Clinical Modification; National Center for Health Statistics (NCHS) 1996–2009*) and DDS-specific codes of full syndrome, residual or suspected autism. Children under 3 years of age entering the DDS system are recorded in the Early Start Report (ESR), which may list autism.

For this analysis, a case is a child with any one of the following: a CDER record with a DDS autism code of 1 (full autism), or an ICD-9-CM code of 299.0 or an ESR record with autism noted or an ICD-9-CM code of 299.0 recorded through February 2006, with or without comorbidities. Children are considered a case based on their earliest record with one of these diagnoses.

A total of 12,125 full syndrome autism cases were identified from the cumulative CDER and ESR records. A computerized search (using SAS 9.1 (SAS Institute, Cary, NC)) was conducted to match them to their California birth record using child and parental names, dates of birth, and social security numbers obtained from the Client Master File maintained by DDS. Questionable matches were reviewed by hand yielding 10,454 cases matched to birth records. 1,447 listed a birthplace outside of California or were not successfully matched to a California birth record and were excluded from further analysis. Of the matched cases, 9,900 were geocoded successfully for a 94 percent success rate among DDS cases listing a California birthplace. Included in the 9,900 cases are 856 from ESR records, 89 percent of which also have a CDER diagnosis of full autism (Table 1). In the full cohort, 2,453,717 (94 percent) live births were geocoded successfully.

Overall incidence was 40 per 10,000 in the five-year birth cohort followed to February 2006. Among these cases, 4.5 percent (n=451) were multiple births; only 40 were higher order than twins.

Spatial analysis methods

As noted above, the 21 RCs are independent organizations with differing diagnostic practices during the study period. For this reason, direct comparisons across RC regions may be invalid; hence each RC region was analyzed separately to identify clusters. Births were assigned to an RC region based on the geocoded mother's address from the birth record. The

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

location of interest for this analysis is the place of exposure, mother's home residence at the time of delivery.

Composite spatial test—To improve overall reliability (estimation of the true risk ratio, sensitivity and specificity) of spatial clustering analysis, we developed a method that applies a set of acceptance criteria to application of three cluster detection tests (CDTs) and one global clustering test (Van Meter et al., 2008). Each of these tests was applied to two sets of areal units. Additionally, one CDT was applied to the data in point form, resulting in seven CDT results and two Global test results for each RC region.

We used three different CDTs, SaTScan (Kulldorff, 1997), FleXScan (Tango & Takahashi, 2005), and Episcan (Christiansen et al., 2006) and one global test, the Maximized Excess Event Test (MEET) (Tango, 2000) because all spatial tests have differing power depending on the shape, size and relative risk of the true underlying cluster. Global tests indicate overall spatial correlation of cases; they are more powerful at detecting the existence of multiple clusters than CDTs but don't identify their locations.

Based on these seven CDT applications, we then defined two categories of clusters: the more stringent Consensus Cluster where results of all seven applications of the CDTs met the qualifying criteria and a less stringent Potential Cluster where CDT results from the point CDT and the three applications in one set of areal units met all qualifying criteria. More than one cluster could be defined in a study region.

The three criteria for a qualifying positive CDT result were: (a) p-value ≤ 0.05 and (b) risk ratio (RR) ≥ 1.7 for an area within an RC region; and (c) the area includes at least 1,000 births. Based on simulations of scan tests using multiple scenarios, we selected the minimum of 1.7 for the observed risk ratio as this was the value of the underlying true RR at which test sensitivity improved dramatically. The p value maximum was the value corresponding to this same underlying RR value, which is also where the accuracy of the observed RR estimation of the true RR greatly increased. The minimum population of 1,000 for a cluster was to ensure that identified clusters were sufficiently robust that they would not disappear if one or two births over the five-year period had been at different locations.

Spatial units used in testing—Geographic data can either be represented as distinct points, identifying each birth location, or aggregated into areal units, identified by a single centroid point and the number of cases and total births within that areal unit. Only one of the CDTs, SaTScan, could handle our large sample size in point form. We therefore created areal units by partitioning each RC region using the Epiunits "Spatial then Density" method (Christiansen & Van Meter, unpublished) to avoid small population units with extreme incidence estimates and wide confidence intervals.

Each RC region was partitioned twice creating two sets of aggregated areal units with population maxima of 1,000 (Set 1) and 2,000 (Set 2) per unit. The two different partition sets allowed two attempts to identify smaller clusters that might in either set be intersected by the areal boundaries, dividing a cluster among several areal units, each with insufficient case numbers to show elevated autism incidence. All three CDTs and the global test were

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

applied to both Set 1 and 2 units for each RC; these, plus the one point location test, yielded a total of seven CDT results along with two results for global clustering per RC region.

When a region's birth density and total population were so low that fewer than 30 areal units of 2,000 maximum births were created (Far Northern and Redwood), a set of partitions with 500 maximum births was substituted.

Analysis of confounding

To assess the contribution of demographic variables to the spatial clustering of autism, we first confirmed their associations with autism in our cohort, and then evaluated their association with the spatial clusters. Association with clustering was focused on the eight RC regions with identified Consensus Clusters, hereafter referred to as the 8-RC study area. The four demographic factors of interest were: mother's age, father's age, highest parental education level, and race/ethnicity.

Because of the earlier noted variability in application of diagnostic criteria across RCs, we included RC of birth as a random effect in mixed regression models.

We first tested each factor as a predictor of autism for the entire cohort, both in bivariate and multiple mixed effects logistic regression models, the latter including all covariates under investigation. These analyses were conducted using SAS 9.1 (SAS Institute, Cary, NC).

Father's age was missing for 6.66 percent of cohort births and 5.43 percent of all autism cases. In the 8-RC study area, these percentages were 5.91 and 4.52, respectively. Removing father's age to permit analysis of a larger cohort had little effect on the regression fit but increased the odds ratio for mother's age by 42 percent. Since the missing proportion is low for both the cases and the full birth cohort, we retained the father's age variable, removing 6.2 percent of observations from the analysis of sociodemographic factors.

Due to their highly significant associations with autism, all four variables were then assessed for associations with spatial clusters. Bivariate mixed Poisson regressions were performed on the combined populations of the 8-RC study area, contrasting births within the boundaries of the ten Consensus Clusters, regardless of case status, to births outside the boundaries of these clusters, but within the 8-RC study area. All demographic variables were significant ($p<0.001$) and retained for the final equation. As the parental age-autism relationship was approximately linear for the log of the rate no higher order terms were used.

We then fit a multiple mixed effects Poisson regression model based on the 8-RC study area by backwards stepwise testing of the four demographic variables, in which variables were retained if the coefficient for at least one level of the variable had $p<0.05$, and if the Akaike information criterion (AIC) indicated improved fit. All four variables were retained in the final equation giving rate ratios for the association of each demographic factor, adjusted for the other covariates, with geographic clusters of autism in California. We thereby compared the distribution of demographic factors for all births in the clusters (regardless of autism status) to the distribution of those factors in all births outside of the clusters.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

From results of fitting the model to the birth population of each RC region, we examined variation across RCs in how strongly the covariates associate with births to mothers residing inside vs. outside Consensus and Potential Clusters. Study area analysis was conducted with R 2.4.0 (The R Foundation for Statistical Computing, www.r-project.org).

For all cases within each cluster, a match was done on each parent to count the number of full and half sibling cases within that cluster. Additionally, multiple births (cases from the same pregnancy) were noted within the sibling cases of each cluster. Relationships more distant than a shared parent could not be explored using our birth record data. Since our study cohort of five years of California births likely does not include all siblings of cases or even all case siblings, no statistical analysis of familial relationships was attempted.

Results

Spatial analysis

Figure 1 indicates the level of clustering determined for each RC region. There are ten Consensus Clusters in eight RC regions and two RC regions each with one Potential Clusters. Redwood Coast, North Bay, San Gabriel/Pomona and East Los Angeles demonstrate no clustering beyond that expected by random processes in the underlying population. Lanterman, Inland and East Bay had significant MEET results, with either one or no qualifying results from any of the CDTs. The results for these last three RC regions indicate global clustering, or spatial correlation of cases, where autism cases did not occur completely randomly but in multiple small clusters; none of the clusters was large enough to be considered a significant cluster on its own. Overall, the clusters our tests identified contained 4.5 to 15.3 percent of RC region births including 9.6 to 24.4 percent of the RC's autism cases. (Appendix Table 1 presents results of the separate spatial tests.)

Table 2 summarizes results for the eight RC regions with Consensus Clusters and two RC regions with Potential Clusters (Figure 2). Of these, the analyses identified two Consensus Clusters in each of the Golden Gate and North Los Angeles RC catchment areas; the six others had one cluster each. Golden Gate, San Diego, and Valley Mountain had non-significant global test results (alpha > 0.077), indicating the improbability of additional clusters beyond those defined.

For Far Northern, Alta, Tri-Counties and Kern, the point-based CDT gave no qualifying clusters, so they fail our composite CDT. However, each had multiple qualifying clusters identified by areal CDT tests. All but Kern also had significant global clustering. Hence, we classified their CDT results as equivocal.

Negative results for three RC regions, two of which fell into the equivocal category, may be due to their very low birth population densities. Redwood Coast, Far Northern and Kern RCs had birth densities of 1.4, 0.9 and 2.5/sq mi, respectively during the entire 5-year study period. Kern RC's very unusual shape, approximating a tilted hourglass, probably further reduced the power of spatial tests.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Demographic analysis

Within each cluster identified, the full sibling cases comprised less than ten percent, and there are no half-sibling cases. Of the nineteen sibling case pairs within clusters, nine were twins. It is unknown whether they were mono- or dizygotic. Although sibships of cases can occur as a result of shared genetics, shared environment, or chance, the limited number of sets of case siblings within the clusters (Table 2) demonstrates that the contribution of inherited susceptibility genes to the observed clustering is not large.

The demographic analysis of the 8-RC study area showed autism clusters to be highly associated with the education of the parent population (Table 3). For six of the ten RCs, the adjusted rate ratio for residence at the time of delivery in the geographic area of an autism cluster, comparing a non-high school-graduate with a college-graduate parent is less than one-fourth. In three of these RCs, the adjusted rate ratio of being born in the cluster area when one or both parents have *some* college education compared with one or both being college graduates was less than half.

The rate ratios for a ten-year increment in parental ages are 1.09 for mothers and 1.05 for fathers comparing births within the clusters of high autism incidence vs. outside those areas. For twenty-year increments (i.e., comparing 40-year old with 20-year old parents), the figures are 1.20 for mothers and 1.10 for fathers.

Births to parents of Hispanic ethnicity were under-represented in each of the ten Consensus Clusters, whereas births to non-white non-Hispanic parents were more common in some RC clusters and less common in others (Appendix Table 2). Parental education was consistently higher in all clusters, and in most RCs, there was a monotonic trend by level of education.

Among individual clusters, there is some variation from the overall pattern in parental ages and race/ethnicity, but none of the effects approach the magnitude of parental education.

Discussion

We undertook a search for areas of increased autism cumulative incidence without hypothesizing a specific exposure, following scientific indications of the possibility of environmental risk factors for autism and public concern over locally increased autism incidence. The method we used was designed to find areas of geographically non-randomly distributed cases of autism. In light of the statistical rarity of autism, we applied a comprehensive cluster detection approach that improves specificity and effect estimation while preserving sensitivity. The resulting spatial analysis defined ten Consensus Clusters and two Potential Clusters of elevated autism risk in California. The majority of these autism clusters were strongly associated with higher education of the parents, a demographic factor previously documented to be associated with increased autism diagnoses.

Finding clusters highly associated with previously identified high-risk demographic groups is useful in assessing the effectiveness of the first application of this statistical technique. The identification of some clusters that were explained by a known risk factor indicates that this procedure can define the location of actual clusters of a rare disorder.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We identified two contiguous clusters covering the boundary between two RC regions: Westside and Northern LA. These two form one large cluster. Thus, analyzing each RC region separately to avoid case definition bias did not preclude identification of clusters that went beyond RC region boundaries.

Once these clusters are located, only those not explained by demographic risk factors are of interest for further exploration of possible environmental exposures from localized or point sources. Our results indicated that clusters of a localized form were mostly explained by demographic risk factors. The findings from this study do not preclude a role for environmental exposures that cluster around nonpoint sources, such as traffic, or that are not clustered spatially because they are widely distributed, such as household products.

Adjusting for other covariates, the rate ratios for births to be located in Consensus Clusters for college-graduate and for white, non-Hispanic parents are higher than are the adjusted autism odds ratios for any of the four demographic study variables in the cohort or in the 8-RC study area. To the extent that demographic factors provide clues to geographic patterns of incidence, this study reinforces the caveat that spatial analysis results should always be assessed with respect to the spatial distribution of the demographic characteristics of the study population.

Spatial Methodologic Issues

There was a group of RCs that had inconsistent CDT results: the equivocal group. The presence of this group emphasizes the need to perform multiple tests, as there was no pattern as to which areal CDTs produced qualifying results. In Tri-Counties RC, FleXScan identified a qualifying result in both Set 1 and 2, but it was a different area in each set.

Of special concern were Kern and Far Northern, where the point-based test did not agree with the areal tests. If the point-based test had not failed, Kern would have a Consensus Cluster and Far Northern would have a Potential Cluster. Spatial tests that are adjusted for population density are more powerful in urban than rural areas (Gregorio et al., 2005). Still, the lack of qualifying results from the point-based SaTScan in low-density RC regions, when all areal-based tests passed our qualification criteria, was unexpected. Gregorio et al. (2005) found general agreement between a point version of SaTScan and one applied to censusbased areal units, which are much larger aggregation units than ours, and had a wider range of unit populations.

Our results suggest the need for a simulation study of CDT performance on rare events comparing analyses using point data with those using aggregated data. Such a study would use various CDTs applied over a wide range of population densities to examine whether tests based on areal units are generally more powerful than ones based on point data, whether they tend to generate false positive results, and what factors influence the balance between sensitivity and specificity. As spatial analysis becomes more common, guidelines for the influence of population density effects, incidence level and data form (point, range of population in aggregated units) on the sensitivity and specificity of CDTs are needed.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Demographic analysis

Since we analyzed children's location at birth, not at diagnosis, it is highly unlikely that parents had moved to a location near an autism treatment center in anticipation of a child's later diagnosis. However, cases in our study met eligibility requirements after parental initiative. DDS utilization could be affected by access to diagnosticians and service providers, as well as knowledge of the RC. Local awareness of treatment options may be higher near specialty autism research and treatment centers. Thus, treatment centers may have an effect on the spatial distribution of autism diagnoses recorded by the DDS. For example, the UCLA Neuropsychiatric Institute and Lovaas Headquarters are located in the vicinity of two RCs with clustering and high overall incidence: North Los Angeles County and Westside RCs. The Central Valley Autism Project (CVAP), Inc. is in Modesto, site of the Valley Mountain RC Potential Cluster. The Northern LA and Valley Mountain clusters were the least associated with parental education.

Any errors related to addresses, e.g., due to non-local moves close to the time of delivery, would be expected to dilute associations and hence the true clustering of autism may be stronger than observed. Additionally, our cases came from an administrative database where the standard of diagnostic accuracy for eligibility determination is not the same as it is for research.

We lacked access to additional sources for cases, e.g. public school records, to augment the DDS cases. Certainly the estimated twenty percent of autism cases not included in the DDS records could have a spatial or demographic bias. If there was an autism treatment center that did not encourage its clients to apply to the DDS system, it would create a local zone of apparently low incidence in our study. Our case definition was dictated by the DDS eligibility criteria. Since this study was looking for clusters that might be associated with point source environmental exposures, it is not likely that including a higher proportion of individuals with milder ASDs as cases would have affected the results.

In our search for raised cumulative incidence of autism using DDS records, most of the areas we identified were highly associated with elevated parental education. There is mounting evidence (Yeargin-Allsopp et al., 2003) that at least some of this clustering by parental education results from the greater access and utilization of services by those with more years of schooling, given that the DDS system relies on parents to voluntarily seek services. It is unknown whether DDS participation among families with an affected child differs by socioeconomic status, race or ethnicity, although services are available regardless of race, ethnicity, wealth, or citizenship. It remains possible that the associated demographic characteristics are surrogates for some other yet-to-be defined/confirmed risk factors, such as subfertility, accumulated exposures, genetic susceptibility or access to optional medical interventions like assisted reproduction or scheduled Caesarean sections.

An association of higher parental education with autism has been shown in recent population-based studies: in the US (Yeargin-Allsopp et al., 2003), (Bhasin & Schendel, 2007) and in the UK (Baird et al., 2006) but not in Denmark (Larsson et al., 2005). The first three studies were conducted in the US and the UK, where population screening is not routine; as in our study population, children are more likely to receive a diagnosis of autism

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

if their parents are more educated. That this association was not found in Denmark where the entire population of three-year-olds is screened is therefore noteworthy.

Acknowledgments

We are grateful for the assistance of Ron Huff, PhD of Alta Regional Center, Paul Choate of the DDS Data Extraction, and the staff of the DDS and the Association of Regional Center Agencies. This study was funded in part by NIEHS P01 grant #ES11269, NIEHS R01 #ES015359, EPA STAR grant #R829388, the UC Davis M.I.N.D. Institute and the UC Davis Center for Animal Disease Modeling and Surveillance (CADMS).

Financial support:

Non-specific support:

UC Davis M.I.N.D. Institute

UC Davis Center for Animal Disease Modeling and Surveillance (CADMS).

Abbreviations

CDER	Client Development Evaluation Report
CDT	Cluster Detection Test
DDS	California Department of Developmental Services
ESR	Early Start Report
ICD-9-CM	International Classification of Diseases, 9th Revision, Clinical Modification; U.S. Department of Health and Human Services
MEET	Maximized Excess Event Test, a global clustering test
RC	Regional Center of the DDS
RR	Relative Risk, incidence within a cluster population/ incidence in the population in the rest of the study region as determined by CDTs

Literature Cited

Akaike H. Data analysis by statistical models]. No To Hattatsu. 1992; 24(2):127–133. [PubMed: 1567644]

Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, et al. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). Lancet. 2006; 368(9531):210–215. [PubMed: 16844490]

Bhasin TK, Schendel D. Sociodemographic risk factors for autism in a US metropolitan area. J Autism Dev Disord. 2007; 37(4):667–677. [PubMed: 16951989]

Center for Health Statistics. Confidential Birth 980-Byte File. State of California Dept of Health Services; 1996–2000.

Christiansen L, Andersen J, Wegener H, Madsen H. Spatial Scan Statistics Using Elliptic Windows. Journal of Agricultural, Biological and Environmental Statistics. 2006; 11(4):411–424.

Croen LA, Grether JK, Hoogstrate J, Selvin S. The changing prevalence of autism in California. J Autism Dev Disord. 2002; 32(3):207–215. [PubMed: 12108622]

Croen LA, Grether JK, Selvin S. Descriptive epidemiology of autism in a California population: who is at risk? J Autism Dev Disord. 2002; 32(3):217–224. [PubMed: 12108623]

Croen LA, Najjar DV, Fireman B, Grether JK. Maternal and paternal age and risk of autism spectrum disorders. Arch Pediatr Adolesc Med. 2007; 161(4):334–340. [PubMed: 17404129]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

DiGiuseppe DL, Aron DC, Ranbom L, Harper DL, Rosenthal GE. Reliability of Birth Certificate Data: A Multi-Hospital Comparison to Medical Records Information. Matern Child Health J. 2002; 6(3): 169–179. [PubMed: 12236664]

Glasson EJ, Bower C, Petterson B, de Klerk N, Chaney G, Hallmayer JF. Perinatal factors and the development of autism: a population study. Arch Gen Psychiatry. 2004; 61(6):618–627. [PubMed: 15184241]

Gregorio D, DeChello L, Samociuk H, Kulldorff M. Lumping or splitting: seeking the preferred areal unit for health geography studies. Int J Health Geogr. 2005; 4(1):6. [PubMed: 15788100]

International Classification of Diseases, 9th Revision. Retrieved 20 June 2007, from ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/ICD-9

Juul-Dam N, Townsend J, Courchesne E. Prenatal, perinatal, and neonatal factors in autism, pervasive developmental disorder-not otherwise specified, and the general population. Pediatrics. 2001; 107(4):E63. [PubMed: 11335784]

Kulldorff M. A spatial scan statistic. Communications in Statistics-Theory and Methods. 1997; 26(6): 1481–1496.

Kulldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer clusters in the northeast United States: a geographic analysis. Am J Epidemiol. 1997; 146(2):161–170. [PubMed: 9230778]

Larsson HJ, Eaton WW, Madsen KM, Vestergaard M, Olesen AV, Agerbo E, et al. Risk factors for autism: perinatal factors, parental psychiatric history, and socioeconomic status. Am J Epidemiol. 2005; 161(10):916–925. discussion;926–918. [PubMed: 15870155]

Lauritsen MB, Pedersen CB, Mortensen PB. Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. J Child Psychol Psychiatry. 2005; (9):46. 963–1071.

Palmer RF, Blanchard S, Stein Z, Mandell D, Miller C. Environmental mercury release, special education rates, and autism disorder: an ecological study of Texas. Health & Place. 2006; 12(20): 203–409. [PubMed: 16338635]

Reichenberg A, Gross R, Weiser M, Bresnahan M, Silverman J, Harlap S, Rabinowitz J, Shulman C, Malaspina D, Lubin G, Knobler HY, Davidson M, Susser E. Advancing paternal age and autism. Arch Gen Psychiatry. 2006; 63(9):1026–1232. [PubMed: 16953005]

Roberts EM, English PB, Grether JK, Windham GC, Somberg L, Wolff C. Maternal Residence Near Agricultural Pesticide Applications and Autism Spectrum Disorders among Children in the California Central Valley. Environ Health Perspect. 2007; 115(10):1482–1489. [PubMed: 17938740]

Roohan PJ, Josberger RE, Acar J, Dabir P, Feder HM, Gagliano PJ. Validation of Birth Certificate Data in New York State. J Community Health. 2003; 28(5):335–346. [PubMed: 14535599]

Tango T. A test for spatial disease clustering adjusted for multiple testing. Stat Med. 2000; 19(2):191–204. [PubMed: 10641024]

Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. Int J Health Geogr. 2005; 4:11. [PubMed: 15904524]

Van Meter KC, Christiansen LE, Hertz-Picciotto I, Azari R, Carpenter TE. A procedure to characterize geographic distributions of rare disorders in cohorts. Int J Health Geogr. 2008; 7:26. [PubMed: 18507863]

Waller LA, Hill EG, Rudd RA. The geography of power: statistical performance of tests of clusters and clustering in heterogeneous populations. Stat Med. 2006; 25(5):853–865. [PubMed: 16453372]

Windham GC, Zhang L, Gunier R, Croen LA, Grether JK. Autism spectrum disorders in relation to distribution of hazardous air pollutants in the san francisco bay area. Environ Health Perspect. 2006; 114(9):1438–1444. [PubMed: 16966102]

Yeargin-Allsopp M, Rice C, Karapurkar T, Doernberg N, Boyle C, Murphy C. Prevalence of autism in a US metropolitan area. Jama. 2003; 289(1):49–55. [PubMed: 12503976]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Appendix

Table 1

Summary results of the parallel spatial testing procedure for clustering of cumulative autism incidence among California births, 1996–2000.

DDS Regional Center	RC #	Type of Clustering	Cluster Detection Test Results														Global MEET Results	
			Point locations				Set 1 areal units w/ 1,000 max births/unit				Set 2 areal units w/ 2,000 max births/unit							
			SaTScan		FleXScan		Episcan		SaTScan		FleXScan		Episcan		SaTScan		Set 1	Set 2
			p	RR[c]	p	RR	p	RR	p	RR	p	RR	p	RR[c]	p	RR	p	p
Lanterman	360	Global	0.04	7/49	0.05	1.71	0.26	2.30	0.18	1.36	0.07	1.54	0.19	1.45	0.16	1.40	0.01	0.02
Golden Gate	361	Consensus	0.02	3.73	0.008	2.98	0.01	2.94	0.008	2.57	0.02	3.80	0.02	4.55	0.01	2.69	0.12	0.14
San Diego	362	Consensus	0.001	2.74	0.001	2.94	0.000	2.22	0.000	2.27	0.001	2.15	0.001	1.96	0.000	1.91	0.26	0.29
Far Northern[a]	363	Global[b]	0.15	2/2	0.001	3.03	0.02	2.43	0.05	2.25	0.01	3.21	0.06	2.90	0.06	3.10	0.01	0.01
Alta	364	Global[b]	0.10	5/69	0.17	2.43	0.05	2.39	0.08	1.74	0.03	1.97	0.03	2.82	0.07	2.06	0.003	0.002
San Andreas	365	Consensus	0.002	2.15	0.003	2.46	0.009	2.16	0.001	1.93	0.001	1.97	0.004	2.18	0.000	1.78	0.001	0.001
Tri-Counties	366	Global[b]	0.01	3/3	0.02	2.46	0.01	2.56	0.09	4.39	0.02	1.89	0.17	2.24	0.11	1.55	0.01	0.004
Central Valley	367	Consensus	0.001	4.33	0.003	2.75	0.04	4.77	0.000	2.36	0.001	3.23	0.01	3.49	0.000	2.23	0.01	0.002
Orange County	368	Consensus	0.02	1.88	0.01	2.14	0.03	2.02	0.006	2.35	0.01	1.94	0.004	2.89	0.002	1.96	0.001	0.001
Inland	369	Global	0.01	3/3	0.40	2.43	0.09	2.29	0.12	1.73	0.14	1.97	0.13	2.02	0.18	1.64	0.01	0.002
Redwood Coast[a]	370	None	0.64	3/42	0.15	3.22	0.23	2.13	0.25	3.86	0.11	3.54	0.24	too sm	0.20	6.14	0.07	0.06
North Bay	371	None	0.28	2/2	0.14	2.30	0.83	2.62	0.65	1.80	0.14	1.90	0.31	1.94	0.23	1.94	0.32	0.29
Kern	372	Equivocal	0.17	2/2	0.002	2.86	0.01	2.25	0.01	2.30	0.05	2.11	0.02	1.93	0.04	2.00	0.24	0.58
East Los Angeles	373	None	0.03	4/9	0.04	1.80	0.04	1.50	0.03	1.48	0.01	1.69	0.04	1.44	0.09	1.90	0.18	0.17
South Central LA	374	Potential	0.05	2.15	0.009	2.48	0.04	1.97	0.01	1.93	0.14	1.83	0.03	1.78	0.01	1.78	0.05	0.02
Harbor	375	Consensus	0.001	1.83	0.002	2.04	0.04	2.12	0.000	1.78	0.001	1.88	0.000	1.72	0.02	1.87	0.001	0.001
Westside	376	Consensus	0.001	1.89	0.001	1.98	0.000	1.74	0.000	1.72	0.002	1.79	0.000	1.74	0.000	1.74	0.001	0.001
Valley Mountain	377	Potential	0.02	3.15	0.04	2.61	0.008	2.71	0.04	3.01	0.20	1.83	0.19	1.63	0.14	1.63	0.08	0.09
North Los Angeles	378	Consensus	0.002	2.18	0.003	2.13	0.009	2.54	0.000	2.07	0.001	1.76	0.000	1.74	0.000	1.82	0.001	0.001
San Gab/Pomona	379	None	0.20	13/550	0.09	2.45	0.07	3.07	0.13	2.20	0.26	2.11	0.09	2.16	0.05	2.16	0.31	0.28
East Bay	380	Global	0.61	2/2	0.37	2.11	0.05	1.66	0.09	1.53	0.04	2.10	0.17	1.50	0.11	2.46	0.001	0.001

[a]Set 2 is replaced by Set 3 with a maximum of 500 births per areal unit

[b]This RC also displayed equivocal CDT results, with half meeting qualifications

[c]#cases / #births in the cluster with the lowest p-value if the only clusters identified with a $p < 0.05$ are smaller than 1,000 births

Table 2

Rate ratios (RRs) of birth location being inside each autism Consensus or Potential Cluster versus not being in a cluster, within the RC region, regardless of autism case status.

Cluster	Education level[a]			Father's Age	Mother's Age	Race/Ethnicity[b]	
	Some College	High School Graduate	Less than HS Graduate	per year older	per year older	Hispanic	Other
	RR[c] (95% CI)	RR[c] (95% CI)	RR[c] (95% CI)	RR[c] (95% CI)	RR[c] (95% CI)	RR[c] (95% CI)	RR[c] (95% CI)
Golden Gate (North)	0.74 (0.66, 0.82)	0.55 (0.48, 0.62)	0.36 (0.29, 0.43)	1.00 (0.99, 1.01)	1.02 (1.01, 1.03)	0.54 (0.47, 0.62)	1.11 (1.03, 1.20)
Golden Gate (South)	0.83 (0.76, 0.90)	0.64 (0.57, 0.70)	0.39 (0.33, 0.46)	1.00 (0.99, 1.00)	1.01 (1.01, 1.02)	0.57 (0.51, 0.63)	0.95 (0.89, 1.02)
San Diego	0.36 (0.34, 0.38)	0.22 (0.21, 0.24)	0.14 (0.12, 0.16)	1.01 (1.01, 1.02)	1.02 (1.02, 1.03)	0.45 (0.42, 0.47)	1.51 (1.44, 1.58)
San Andreas	0.65 (0.62, 0.68)	0.55 (0.52, 0.59)	0.34 (0.31, 0.37)	1.00 (0.99, 1.00)	0.99 (0.98, 0.99)	0.73 (0.69, 0.78)	1.61 (1.54, 1.68)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Cluster	Education level[a]			Father's Age	Mother's Age	Race/Ethnicity[b]	
	Some College	High School Graduate	Less than HS Graduate	per year older	per year older	Hispanic	Other
Central Valley	0.57 (0.54, 0.60)	0.43 (0.40, 0.45)	0.24 (0.22, 0.26)	0.99 (0.99, 1.00)	1.00 (1.00, 1.01)	0.59 (0.56, 0.62)	1.13 (1.07, 1.20)
Orange County	0.62 (0.58, 0.65)	0.37 (0.34, 0.39)	0.20 (0.18, 0.22)	1.01 (1.00, 1.01)	1.02 (1.01, 1.02)	0.45 (0.42, 0.47)	0.47 (0.44, 0.49)
South Central LA[d]	0.65 (0.61, 0.69)	0.37 (0.34, 0.39)	0.13 (0.12, 0.14)	1.00 (1.00, 1.01)	1.01 (1.01, 1.02)	0.25 (0.23, 0.27)	0.12 (0.11, 0.13)
Harbor	0.47 (0.44, 0.49)	0.28 (0.26, 0.30)	0.10 (0.08, 0.11)	1.01 (1.01, 1.01)	1.03 (1.02, 1.03)	0.42 (0.40, 0.45)	0.55 (0.52, 0.57)
Westside	0.38 (0.34, 0.41)	0.23 (0.20, 0.25)	0.25 (0.21, 0.28)	1.01 (1.00, 1.01)	1.00 (0.99, 1.00)	0.38 (0.34, 0.41)	0.34 (0.32, 0.36)
Valley Mountain[d]	0.89 (0.81, 0.99)	0.92 (0.83, 1.01)	0.61 (0.53, 0.69)	1.00 (0.99, 1.00)	0.99 (0.98, 1.00)	0.66 (0.61, 0.71)	0.56 (0.51, 0.62)
North LA County (West)	0.65 (0.59, 0.70)	0.62 (0.57, 0.67)	0.69 (0.62, 0.76)	1.01 (1.00, 1.01)	1.01 (1.00, 1.02)	0.65 (0.61, 0.70)	0.62 (0.57, 0.67)
North LA County (Center)	0.66 (0.63, 0.70)	0.61 (0.58, 0.64)	0.62 (0.59, 0.66)	1.01 (1.01, 1.01)	1.01 (1.00, 1.01)	0.64 (0.62, 0.67)	0.63 (0.60, 0.66)

[a]Reference level: College graduate and above

[b]Reference level: White, non-Hispanic

[c]From separate multiple Poisson regression equations for each cluster within its RC region, adjusted for other covariates

[d]Potential Cluster

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

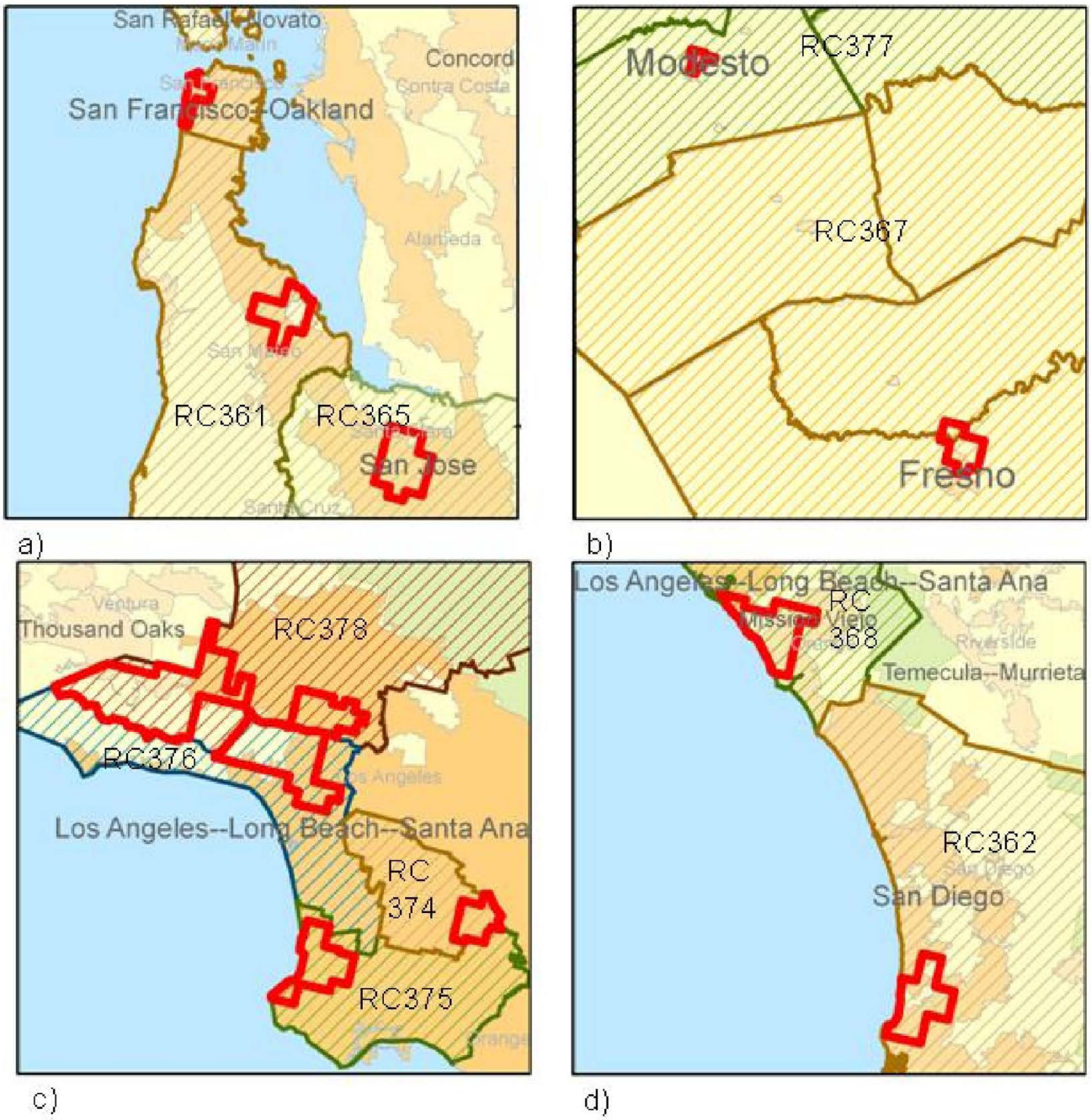

Figure 1.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

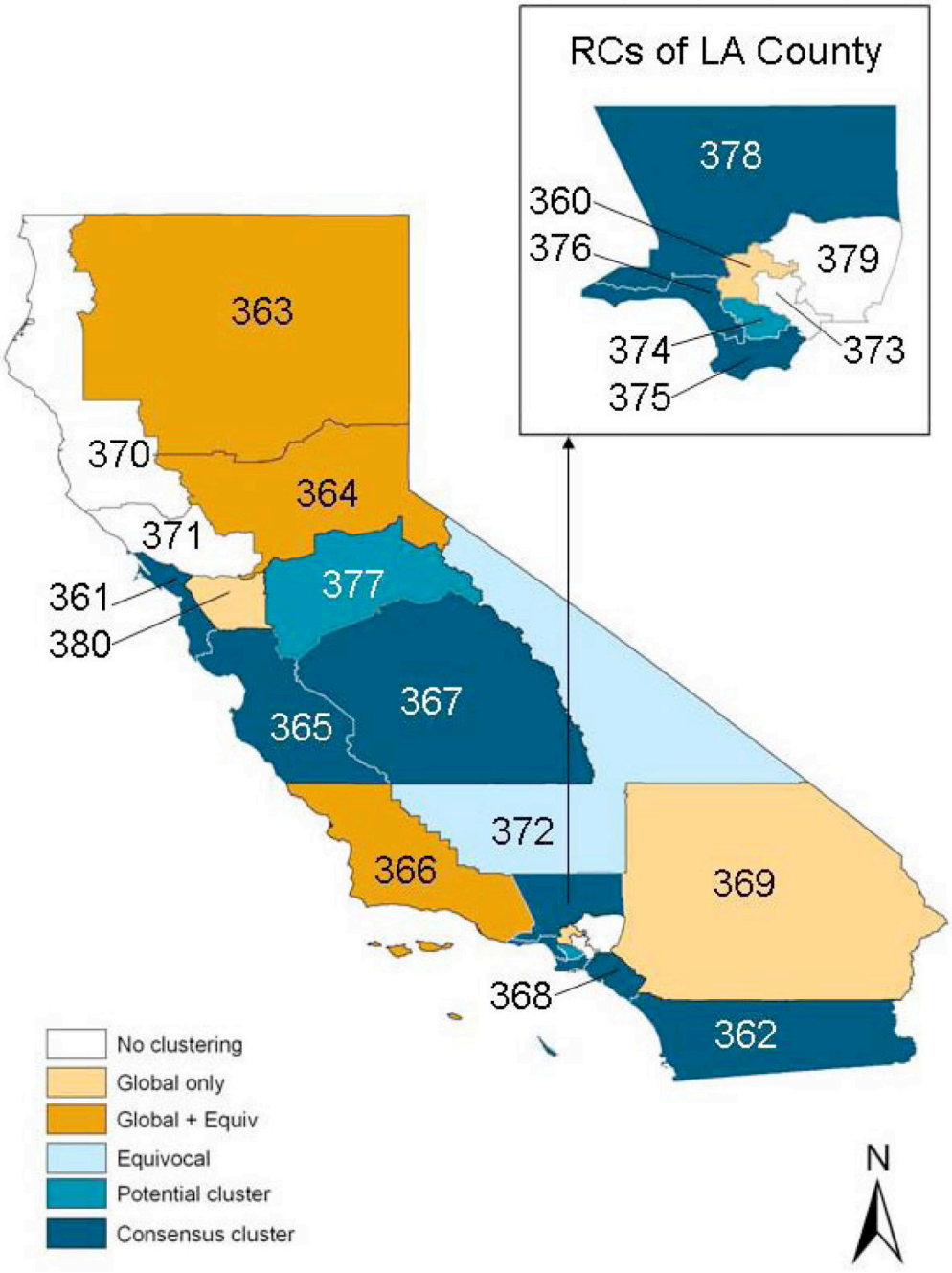

Figure 2.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

The study cohort: All geocoded births in California, 1996–2000, including autism cases from Dept. of Developmental Services cumulative Feb 2006 records.

DDS Regional Center	RC Number	Total Births	Total Autism Cases	Cases per 10,000 Births	# Cases not dx in Birth RC[a]	% Cases not dx in Birth RC[a]
Lanterman	360	92672	816	88.05	217	26.59
Golden Gate	361	101199	253	25.00	83	32.81
San Diego	362	219417	628	28.62	67	10.67
Far Northern	363	27264	95	34.84	7	7.37
Alta	364	127804	333	26.06	39	11.71
San Andreas	365	174980	606	34.63	140	23.10
Tri-Counties	366	92102	385	41.80	46	11.95
Central Valley	367	133702	269	20.12	28	10.41
Orange County	368	219108	837	38.20	139	16.61
Inland	369	237579	576	24.24	84	14.58
Redwood Coast	370	13235	38	28.71	6	15.79
North Bay	371	58979	186	31.54	23	12.37
Kern	372	52973	145	27.37	10	6.90
East Los Angeles	373	95977	674	70.23	113	16.77
South Central LA	374	146768	504	34.34	143	28.37
Harbor	375	108292	675	62.33	114	16.89
Westside	376	69899	604	86.41	144	23.84
Valley Mountain	377	79224	273	34.46	20	7.33
North Los Angeles	378	144550	1020	70.56	98	9.61
San Gab/Pomona	379	98506	453	45.99	102	22.52
East Bay	380	159487	530	33.23	109	20.57
Statewide total		2453717	9900	40.35	1732	17.49

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Consensus and Potential Clusters of increased autism cumulative incidence, by Regional Center of birth.

DDS Regional Center	Range of RR results for CDTs[a]	Cluster births		Cluster autism cases		Cluster cases that are siblings		
		Total births in cluster	As % of RC births	Total Cases In cluster	As % of RC cases	# pairs among cluster cases	As % of cluster cases	# of pairs that are twins
Golden Gate[b]	2.69, 4.55	7853	7.76	49	19.37	2	8.16	2
San Diego	2.22, 2.94	9804	4.47	60	9.55	1	3.33	1
San Andreas	1.93, 2.46	12782	7.30	77	12.71	1	2.60	0
Central Valley	2.36, 4.77	10528	7.87	48	17.84	0	0.00	0
Orange County	1.94, 2.89	12013	5.48	80	9.56	0	0.00	0
South Central LA[c]	1.93, 2.48	9249	6.30	58	11.51	2	5.41	0
Harbor	1.78, 2.12	10733	9.91	111	16.44	3	5.83	1
Westside	1.72, 1.98	6897	9.87	103	17.05	3	4.82	2
Valley Mountain[c]	2.61, 3.01	3834	4.84	32	11.72	1	6.90	1
North LA County[b]	2.07, 2.54	22146	15.32	249	24.41	6	6.25	2

[a]Relative risk of being a DDS autism case born within versus outside the clusters

[b]Two Consensus Clusters

[c]Potential Cluster

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Demographic contributions to geographic clustering of autism among California births, 1996–2000, based on a multiple mixed Poisson regression.

Covariate	Adjusted rate ratio[a] (95% CI)
Parental maximum education level[b]	
Some College	0.55 (0.54, 0.56)
High School Graduate	0.42 (0.41, 0.43)
Less than high school graduate	0.33 (0.32, 0.34)
Parental age	
Father's Age, per decade	1.05 (1.04, 1.06)
Mother's Age, per decade	1.09 (1.08, 1.11)
Child's Race/Ethnicity[c]	
Hispanic	0.55 (0.54, 0.56)
Other	0.82 (0.80, 0.83)

[a] Rate ratio of a birth being in vs. outside a Consensus Cluster, regardless of case status, within the 8-RC study area, adjusted for other covariates and RC differences

[b] Reference level: College graduate and above

[c] Reference level: white, non-Hispanic