Spatial clusters of autism births and diagnoses point to contextual drivers of increased prevalence

Soumya Mazumdar[a], **Alix Winter**[b], **Ka-Yuet Liu**[c], and **Peter Bearman**[d]
[a]Australian National University, Canberra, AU

[b]Harvard University, Cambridge, MA, US

[c]University of California, LA, US

[d]Columbia University, NY, US

Abstract

Autism prevalence has risen dramatically over the past two decades in California. Although often suggested to have been crucial to the rise of autism, environmental and social contextual drivers of diagnosis have not been extensively examined. Identifying the spatial patterning of autism cases at birth and at diagnosis can help clarify which contextual drivers are affecting autism's rising prevalence. Children with autism not co-morbid with mental retardation served by the California Department of Developmental Services during the period 1992 to 2005 were matched to California's Birth Master Files. We search for spatial clusters of autism at time of birth and at time of diagnosis using a spatial scan approach that controls for key individual-level risk factors. We then test whether indicators of neighborhood-level diagnostic resources are associated with the diagnostic clusters and assess the extent of clustering by autism symptom severity through a multivariate scan. Finally, we test whether children who move into neighborhoods with higher levels of resources are more likely to receive an autism diagnosis relative to those who do not move with regard to resources. Significant birth and diagnostic clusters of autism are observed independent of key individual-level risk factors. While the clusters overlap, there is a strong positive association between the diagnostic clusters and neighborhood-level diagnostic resources. In addition, children with autism who are higher functioning are more likely to be diagnosed within a cluster than children with autism who are lower functioning. Most importantly, children who move into a neighborhood with more diagnostic resources than their previous residence are more likely to subsequently receive an autism diagnosis than children whose neighborhood resources do not change. We identify birth and diagnostic clusters of autism in California that are independent of individual-level autism risk factors. Our findings implicate a causal relationship between neighborhood-level diagnostic resources and spatial patterns of autism incidence but do not rule out the possibility that environmental toxicants have also contributed to autism risk.

Keywords

autism; spatial clustering; California; neighborhood resources; GIS; geography; mobility

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 prevalence has risen dramatically over the past two decades (Mitka, 2010). In California, the increase has been precipitous – over 600% from the 1992 to 2002 birth cohorts (Keyes, Susser, Cheslack-Postava, Fountain, Liu, & Bearman, 2011). Similar rates of increase have been observed elsewhere in the United States and in other industrialized countries (Baio, 2002; Madsen, Lauritsen, Pedersen, Thorsen, Plesner, Andersen et al., 2003). Numerous biological, environmental and social factors have been implicated in the rise of autism, but there is no general consensus as to the roles that each have played. The majority of empirical studies have focused on potential genetic causes of autism (Abrahams & Geschwind, 2008) and other individual level risk factors, such as male sex, advanced parental age, prenatal and perinatal complications, and maternal exposures to viruses and other teratogens (Kolevzon, Gross, & Reichenberg, 2007; Patterson, 2009). Although often suggested to have been crucial to the rise of autism, environmental and social contextual drivers of diagnosis, such as local environmental toxicants (Windham, Zhang, Gunier, Croen, & Grether, 2006), diagnostic accretion (King & Bearman, 2009), legislative change (Fountain & Bearman, 2011), neighborhood level resources (King & Bearman, 2011) and increased awareness (Liu, King, & Bearman, 2010), have been studied less extensively.

This article utilizes administrative data from California to consider how identifying the spatial patterning of autism cases at birth and at diagnosis can inform the study of contextual drivers of autism. We have previously identified a spatial cluster of autism cases at birth (henceforth "birth cluster") located in the West Hollywood (Mazumdar, King, Liu, Zerubavel, & Bearman, 2010). Given that not all families live at the same residences from the time of their children's births to diagnoses, examining the spatial clustering of autism at these two moments in time can help disentangle the contextual mechanisms involved. For example, it is possible for a birth cluster to form as a result of the neighborhood-level clustering of an autism risk factor that is particularly relevant around the time of birth, such as an environmental toxicant. Meanwhile, a different set of mechanisms that are independent of those generating birth clusters could be responsible for clustering at diagnosis. Such "diagnostic clusters" could be observed if parents who suspect that their children may have autism select neighborhoods based on available services or neighborhoods that parents select to move to are associated with an increased risk of acquiring an autism diagnosis.

In California, salient neighborhood-level characteristics that could be associated with an increased risk of autism diagnosis include: socioeconomic status (SES), pediatrician density, advocacy organization density, and spending by the Department of Developmental Services (DDS). Children residing in high SES neighborhoods are at greater risk of receiving an autism diagnosis (King & Bearman, 2011; Liu et al., 2010), a finding that is consistent with the effect of neighborhood SES on a wide range of other health outcomes. For example, neighborhood SES is a predictor of the stage at which cancer is diagnosed (Breen & Figueroa, 1996; Shipp, Desmond, Accortt, Wilson, Fouad, & Eloubeidi, 2005). The exact mechanisms underlying the associations between neighborhood SES and health outcomes are debated, but a likely component is that neighborhood SES is a good proxy for local resources and the availability of health-related information (Eng, Maxfield, Patrick, Deering, Ratzan, & Gustafson, 1998). Meanwhile, the number of pediatricians in a neighborhood provides a more specific measure of available resources. There is extensive literature that shows that physician density (Ananthakrishnan, Hoffmann, & Saeian, 2010; Léonard, Stordeur, & Roberfroid, 2009; Roll, 2012) is associated both with the timely diagnosis of certain disorders and with increased consumption of medical services (Menken & Sheps, 1985). While a sufficient supply of pediatricians may be able to address the need for diagnostic services, advocacy organizations are key to spreading awareness of symptoms among caregivers. An increased density of advocacy organizations should, therefore, be

associated with an increased likelihood of autism diagnosis in a given neighborhood. In addition, advocacy organizations can influence legislation and funding. Lastly, California's DDS coordinates autism services through a network of 22 regional centers. Residents of California are assigned to regional centers by zip code, and services are available to children diagnosed with autism and other mental disorders free of charge. It has been argued that the availability of free services may encourage parents whose children exhibit mild autism symptoms to actively pursue a diagnosis (Zarembo, 2011). Yet, variations in regional center funding may increase or decrease this incentive differentially by area. Together, the above characteristics capture neighborhood resources from different perspectives associated with access to diagnostic services and awareness of diagnostic symptoms.

Road map

These potential relationships between neighborhood-level characteristics and autism have specific implications for the spatial patterning of autism incidence. First, whether due to migration, to having been generated by different sets of contextual mechanisms, or a combination of both, birth and diagnostic clusters are unlikely to completely overlap. Therefore, in this study we first identify birth and diagnostic clusters of autism in California and then assess their overlap. It is possible that parents who are at greater risk of having children with autism live in the same neighborhoods or parents whose children are at greater risk for autism similarly select neighborhoods to move to after their children are born. To address these possibilities of residential sorting, we control for individual-level characteristics of parents when identifying the clusters.

Second, if there is substantial overlap between birth and diagnostic clusters, it will not be possible to empirically distinguish which of the two moments in time is more relevant. Given that a substantial proportion of children (>50%) do not move between the time of birth and the time of diagnosis, the presence of birth clusters could lead to clustering at time of diagnosis. Yet, it is equally likely that the presence of diagnostic clusters caused by mechanisms present at time of diagnosis could lead to the observation of clustering at time of birth. We, therefore, use the following tests to help identify the contextual mechanisms most relevant to the rise of autism.

If mechanisms related to diagnosis are responsible for generating diagnostic clusters, they should be positively associated with level of neighborhood resources. We examine whether the four key neighborhood level resources mentioned above are associated with the diagnostic clusters more so than with autism diagnoses in California in general. Next, there is considerable ambiguity in the diagnosis of autism spectrum disorders and, consequently, physicians' responses to symptom presentation are heterogeneous (Bresnahan, Li, & Susser, 2009; Eyal, 2010; Lecavalier, Snow, & Norris, 2011; Noterdaeme, Wriedt, & Höhne, 2010; Saulnier & Klin, 2007). Even when holding all design and methodological factors invariant, prevalence estimates have varied by a factor of 4.5 from the strictest to the least demanding set of diagnostic criteria (Charman, Pickles, Chandler, Wing, Bryson, Simonoff et al., 2009). Therefore, an increased level of neighborhood-level resources in terms of pediatrician and advocacy organization density, regional center spending, and SES would lead to more diagnoses of *high-functioning* autism. We thus test whether autism diagnoses cluster by severity. Finally, focusing on children who have been exposed to varying levels of diagnostic resources allows one to more clearly assess whether they have had an impact on the rising incidence of autism. If they have, children who moved into neighborhoods with higher levels of resources should have a higher chance of being diagnosed with autism relative to children whose levels of resources did not change. We assess whether children who moved into highly resourced neighborhoods are at significantly higher risk of *subsequent* autism diagnosis than children whose level of resources never changed.

METHODS

Study population

We obtained information on clients with Autistic Disorder (International Classification of Disease-9 299.0) served by the DDS from 1992 to 2005. It has been estimated that 80% of all children with autism in California are served by the DDS. The remaining 20% have other diagnoses on the autism spectrum, such as Asperger's, that do not by themselves qualify an individual for DDS services (Croen, Grether, Hoogstrate, & Selvin, 2002). We further confined our analyses to children with "sole autism," those whose diagnoses are not co-morbid with mental retardation.

Each DDS client is evaluated annually using the Client Development Evaluation Report (CDER) in order to determine appropriate services based on level of functioning. We utilized the average score of three CDER items that relate to communication functioning (word usage, receptive language, and expressive language) from clients' first evaluations as a composite communication score. We categorized this score into four categories: highest functioning, high functioning, mid functioning and lower functioning. Since the functioning scores improve as children age, the cutoff scores for 3 and 4 year olds are specific to their age groups and correspond to the 10th, 20th, and 75th percentiles in the 1992 birth cohort. The percentage of children in the highest functioning category changed from 10% in 1992 to 19% in the 2002 birth cohort (Dakhlallah & Bearman, 2012).

Information on all children born in California along with individual level risk factors was obtained from California's Birth Master Files (BMF). The BMF contains detailed demographic information related to the child, mother, and father as well as prenatal and birth characteristics. We linked the DDS data to the BMF of all children born from 1992 to 2002 using probabilistic matching. Links were based on first, middle, and last names, sex, race, date of birth, and maternal zip code at birth. The resulting linked dataset was manually reviewed. About 80% of children in the DDS database were linked to a birth record (Liu et al., 2010), and the majority of children without a link were born outside of California (King, Fountain, Dakhlallah, & Bearman, 2009). A more detailed description of the linking process can be found in our previous studies (King & Bearman, 2011; Liu et al., 2010; Mazumdar et al., 2010). Information on the following known autism risk factors was obtained from the BMF: male sex, whether the birth was paid for by Medi-Cal (California's Medicaid program), preterm birth (<34 weeks), low birth weight status (<2.5 kg), mean parental age, mean parental education, and race/ethnicity. When information on father's age, education or race was missing, mother's age, education or race was used. Ten percent of records were missing information on father's age, while a negligible number of records (<1%) were missing information on both parents' ages. When information on both parents was missing, the record was dropped. Descriptive statistics on the linked DDS-BMF dataset are provided in Appendix Table 1.

The BMF also contains each child's address at birth, which we geocoded to Zip Code Tabulation Areas (ZCTAs) for the years 1993 to 1996 and to census block groups for the years 1997 to 2002 (98% success rate). ZCTAs are statistical geographic entities produced by the U.S. Census Bureau that are generalized approximations of the United States Postal Service's ZIP Codes. There are approximately 1,600 ZCTAs and 22,000 block groups in California. Using the same method, we also geocoded the case children's addresses at diagnosis, which were obtained from the DDS. Children whose addresses could not be geocoded (<2%) were dropped from the dataset.

Because it is most common to receive an autism diagnosis at ages three or four years (Liu et al., 2010), we restricted our analyses to those who were diagnosed at ages three or four

years. Our final dataset included 8,044 children with sole autism born between 1993 and 2002 and diagnosed between 1997 and 2005. (Data censorship prevented us from including children born in 2002 and diagnosed in 2006.)

Siblings

Since we are interested in the effect of neighborhood-level resources on autism diagnoses for families that move across neighborhoods, it is necessary to know the post–childbirth movement patterns of families, independent of whether their children are later diagnosed with autism. To accomplish this, we constructed a dataset of siblings (includes 1,954,862 children). We matched each child in the BMF born between 1997 and 2007 to their full siblings using parents' dates of births and the first letter of mothers' maiden names. This dataset has been described in more detail elsewhere (Liu et al., 2010). We assumed that families remained intact and travelled together through our study period and were able to infer the locations of older siblings at the times of births of younger siblings.

Since we also have location data from the DDS for children later diagnosed with autism, we know which families moved prior to their children's diagnoses and which families stayed within a ZCTA between birth and diagnosis. This information forms the case data. The control group is comprised of children with a younger sibling in the sibling dataset who did not receive an autism diagnosis either before or after their families moved. To ensure that the case and control populations were comparable except for diagnostic status, we matched the case and control populations using exact and propensity score matching. A propensity score is defined as the conditional probability of assignment to a particular treatment versus control group given a set of observed covariates (Leuven & Sianesi, 2003). In our case, the "treatment" group is children who were diagnosed with autism, and the control group is children in the sibling data without an autism diagnosis. By matching our case and control groups using propensity score matching, we minimize the risk of making erroneous inferences regarding the effects of neighborhood-level resources due to confounding effects of between-group differences in any observed characteristics that might also be associated with neighborhood-level resources. One of the advantages of propensity score matching is that the treatment effect identified does not depend on how correctly linear models are specified, such as in the case of covariate adjustments in multivariate regression, which can be difficult to determine. Instead, the appropriateness of the propensity score matching can be easily determined by checking the balance in the covariates (Oakes & Johnson, 2006). We exact matched cases to controls on the children's ages and used propensity scores to match cases and controls within each age stratum using information on parents' ages, sex, Medi-Cal, mother's race and education. We used a caliper of $1/4^{th}$ the standard deviation and matched approximately 10 controls to each case. The matched dataset has 34,693 controls and 3,703 cases. Appendix Table 2 displays the numbers of siblings born in each year for the control population.

Additional Data

In order to calculate risk of autism diagnosis in the population, the numbers of children aged 0 to 4 years in each ZCTA were obtained from the Earth Sciences Research Institute's (ESRI) Community Sourcebook America (ESRI, 2000-2005) for the years 2000 to 2005. For the years 1997 to 1999, for which similar data were not available, we used linear extrapolation to derive the number of children living in each ZCTA.

To measure physician and advocacy organization densities, we obtained data on pediatricians from Medical Marketing Services, which licenses the data from the American Medical Association, and data on advocacy organizations from the Internal Revenue Service (IRS) database of tax-exempt charity organizations. 3,162 pediatricians and approximately

3000 advocacy organizations per year were then geocoded to ZCTAs. In addition, median income data by ZCTA were obtained from ESRI (ESRI, 2000-2005), and regional center spending data were obtained from the DDS and assigned by catchment area.

Cluster Analysis

To identify neighborhoods with excess risk of autism, or "clusters," we used the Kulldorff's Spatial Scan Statistic. Kulldorff's Spatial Scan Statistic reduces cluster detection to maximum likelihood estimation or finds the "most likely cluster" over geographic space. It is implemented with SaTScan Software (Kulldorff, 2006). By conducting just one hypothesis test over the entire geographic space, this method solves the problem of multiple hypothesis testing that has plagued cluster detection literature. SaTScan identifies candidate clusters, circles of increasing radii bound by a maximum threshold radius and centered on pre-specified locations, such as ZCTA centroids. Over many candidate clusters, SaTScan maximizes a likelihood ratio that is a function of the observed number of cases 'O', the expected number of cases in the candidate cluster 'E', and the total number of cases in the entire region (California), '-n.' Since a circle can encompass multiple ZCTAs, E and O can be aggregate statistics. For details on the derivation of the log likelihood ratio and the scanning procedure, (Kulldorff, 1997). This 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 9,999 Monte Carlo randomizations of the data to assess statistical significance. The Poisson Model is appropriate when case and population data are aggregated to counts, such as in our search for neighborhoods of excess autism births relative to all births and neighborhoods of excess diagnoses relative to population controls at the local levels of ZCTAs and block groups. The area that has the highest likelihood value (or the lowest p value) is the most likely primary cluster. Note that if O is less than E in a given cluster, the relative risk would be less than 1, resulting in a region classified as a cluster of low relative risk.

The quantity E represents the expected number of autism cases given the average individual-level risk within the candidate cluster. As mentioned above, residential sorting of parents with an increased number of autism risk factors into certain neighborhoods could result in an uneven spatial distribution of individual risk, potentially masking environmental and social contextual risk factors. Unfortunately, the SaTScan program does not allow for covariate correction for a large number of variables at the individual level (Kleinman, Abrams, Kulldorff, & Platt, 2004). Yet, we can correct the expected numbers of cases of autism at the ZCTA and block group levels given the areas' compositions of individuals using a regression model: for each annual cohort, we estimated the probability of each child born being subsequently diagnosed with autism using the individual autism risk factors mentioned above in a logistic regression model:

$$p_{ij} = e^{\left(\alpha + \Sigma m \, \beta m^* X m + {}^\epsilon m\right)/\left(1 + \alpha + \Sigma m \, \beta m^* X m + {}^\epsilon m\right)} \quad (1)$$

where p_{ij} is the estimated probability for the i'th individual in the j'th local area (ZCTA or block group), and α and βm are the estimated intercepts and slopes for m risk factors. (In our case, m=7.) We approximated the risk of autism diagnosis at each local area as the $Median(p_{ij}) = p_j$ for i=1 to n_j births, and we calculated the expected number of autism cases E_j in j as $p_j{}^*n_j = E_j$. The median provides a more stable estimate of risk than the mean, because a small number of individuals in a local area might have extremely high or low values of p_{ij}. This, in turn, would bias the mean p_{ij} and result in a biased expected count E_j. The median, in contrast, provides a robust summary measure of local area risk level.

The population data used to estimate the p_{ij}'s and n_j's were different for the analyses of births and diagnoses. For calculating the risk of autism births, the BMF were used for all births, and the n_j were the number of births in a block group or ZCTA j. For calculating the risk of autism diagnoses, data on children aged 0 to 4 years were required. Since such data were usually not available, we used the linked sibling dataset to estimate the median p_j's at each ZCTA. The probabilities p_j were then multiplied by the number of 0-4 year olds in each ZCTA(n_j).

We discarded any clusters with E_j's less than 1, since, in theory, these areas had zero expected cases. A Spatial Scan then aggregated the remaining E_j's over circles of increasing radii that were bound by a maximum threshold radius (set here at 1% of the expected population) and centered on pre-specified local area centroids to create the candidate clusters and calculate E, the expected number of cases. We have previously shown that a 1% upper threshold is appropriate for our data (Mazumdar et al., 2010). We mapped the clusters and their temporal stability over time and tabulated the risks associated with them.

To assess the public health impact of these clusters, we calculated their Population Attributable Fractions (PAF) (Yiannakoulias, 2009) as:

$$\left(O_c / Z\right)^* \left(RR_c - 1\right) / RR_c \quad (2)$$

where O_c is the observed number of cases in cluster c, Z is the total number of cases in California over the observation window, and RR_c is the relative risk in cluster c. The PAF indicates the reduction in risk gained by the elimination of an exposure or, in this context, how much of the overall risk of sole autism in California is attributable to the observed clusters (Yiannakoulias, 2009).

Neighborhood resources

Once spatial clusters were identified, simple Geographic Information System (GIS) queries were used to measure the degree of association between neighborhood-level resources and diagnostic clusters adjusted for individual-level risk factors. We calculated the densities of pediatricians and advocacy organizations by ZCTA by dividing the numbers of pediatricians and advocacy organizations by the numbers of 0 to 4 year old children. In addition, median income and regional center spending at the ZCTA level were calculated. A reasonable analysis of the effect of neighborhood-level characteristics on autism diagnoses should also include a measure that is correlated with urbanicity but is unrelated to autism diagnoses. The absence of a relationship between such a variable and autism diagnoses would underscore the robustness of our test. Therefore, we also constructed a dataset of 5000 locational points per year distributed over California and then calculated the densities of these points among 0 to 4 year olds in each ZCTA. Since these points mimic the underlying population distribution, whether a significant relationship is found between their densities and autism diagnoses forms a test for whether any associations found between neighborhood-level resources and diagnostic clusters are simply due to urbanicity.

Spatial clustering by functioning

As argued above, neighborhood diagnostic resources should be particularly relevant to the diagnosis of children with less severe autism symptoms. Therefore, we examined whether children with autism in the highest functioning category are spatially clustered. We used the multivariate scan statistic (Kulldorff, Mostashari , Duczmal, YW, Kleinman, & Platt, 2007) to evaluate each of the functioning categories at once relative to the underlying population at risk. In the multivariate scan, for each candidate cluster, the likelihoods for each category are calculated separately and then summed together. This summed likelihood indicates the

likelihood that a specific neighborhood is a cluster. Therefore, clusters may be made up of either disproportionate contributions from different categories or proportionate contributions from all categories. Thus, the multivariate scan can be used to identify clusters of children with autism who are members of any of the four severity groups. The multivariate scan method is summarized in the context of our autism functioning data in Appendix Table 3 (Kulldorff, 2011; Kulldorff et al., 2007). Using this approach, it is possible for an excess of diagnoses in a neighborhood to consist entirely of children who are in the highest functioning category. Such a result would point toward different mechanisms (e.g. diagnostic expansion) than if the excess were equally attributable to all four categories. Appendix Table 4 summarizes all the spatial scans we used in our analyses.

Neighborhood level resources and mobility

Lastly, we assessed the impact of neighborhood-level resources at time of diagnosis on autism incidence independent of locational factors at time of birth by focusing on families that moved. We used the matched sibling dataset and measured changes in the levels of neighborhood resources by dividing the resource variables – SES, regional center spending, and pediatrician, advocacy organization, and random point density - into deciles. For the density measures, we added an extra category for zero density. Although the categorization of continuous data may result in loss of information, it allowed us to detect non-linear associations. For each resource attribute, the matched dataset was structured as either a 10 by 10 or 11 by 11 origin-destination table by linking children's origin and destination ZCTAs with the ZCTAs' respective neighborhood resource categories. The tables, therefore, captured the movements, or lack thereof, of cases and controls from one resource category to another. We lagged the attribution of the destination ZCTA by two years to ensure temporal and spatial causality and to adjust for the temporal uncertainty that families may have moved in the year previous to the one in which a child was diagnosed (cases) or a sibling was born (control).

For each neighborhood-level attribute, we then modeled the odds of autism diagnosis for a child that moved from one resource category to another relative to a given reference category. We used log linear models, extensions of the two-way standard contingency table, which analyzes the conditional relationship between two or more discrete, categorical variables by taking the natural logarithm of the cell frequencies within a contingency table. Log linear models are especially relevant to studying social (Breen, 2006, 2010) and geographical mobility (Kaldor, Khlat, Parkin, Shiboski, & Steinitz, 1990), where the goal is to model the odds of an outcome resulting from movement across categories in an Origin-Destination Matrix. The log linear model we used is as follows (Marsden, 1988):

$$\log\left(m_i / z_i\right) = \Sigma_j \, \beta_j x_{ij} \quad (3)$$

where is the number of cases of autism in a cell i, z_i is the number of controls in the cell, β_j is a vector of particular log linear parameters to be estimated, and x_{ij} is an element of a pre-specified design matrix. One of our design matrices is provided in Appendix Table 5. It indicates possible movements across the levels of an origin-destination table. Our matrices included D, indicating downward movement (from a higher to lower level of resources) 0, indicating no movement, 1 through 6, indicating increases of 1 through 6 resource deciles respectively, and 7+, indicating an increase of 7 or more resource deciles. We fitted the log linear model to the four 10 by 10 or 11 by 11 resource matrices (pediatrician density, advocacy organization density, median income, and regional center spending) and the control matrix of random point density using LEM software (Vermunt, 1997).

RESULTS

A primary cluster is observed for each birth cohort from 1993 to 2002 after adjustment for known autism risk factors. Figure 1.A displays the clusters by temporal stability. Darker shades indicate longer duration of an area being part of a cluster. The clusters consist of three separate regions of high risk centered on Santa Monica, Alhambra and North Hollywood. The cluster centered on North Hollywood has an approximate radius of 10 kilometers and is bounded by the South Central regional center to the South, the North Los Angeles regional center to the North, and Interstate 5 to the West. Panels B and C show the relative risks and PAFs for each birth cohort's cluster. The risk of a child later diagnosed with autism having been born within a birth cluster relative to the risk of a child later diagnosed with autism having been born elsewhere in California ranges from 3.6 in 1996 to 6.0 in 1993. The PAFs of these clusters vary from 2.6% in 1996 to 4.3% in 1993.

Figure 2 displays the unadjusted diagnostic clusters, and Figure 3 displays the same clusters after adjustment for autism risk factors. They both include the adjusted birth clusters in crosshatch for easy reference. Table 1 shows the relative risks and PAFs for the adjusted and unadjusted diagnostic clusters. Adjusting for known autism risk factors shrinks the geographic size of the clusters and reduces their relative risks. For example, the areas north of Malibu are part of the unadjusted diagnostic clusters but not part of the adjusted clusters. Although the average relative risk over the study period is reduced from 5.4 to 3.7 with adjustment, it is still quite significant. The PAFs exhibit similar decreases.

Approximately 55% of the area of the adjusted diagnostic cluster is also part of the adjusted birth cluster. Net of a possible contribution from modeling errors, this lack of complete overlap implies that contextual mechanisms that are specific to the diagnostic process might have played a role in increased autism incidence. Furthermore, about 47% of children who received their diagnoses while living in a diagnostic cluster were born outside of a birth cluster. This rules out the possibility that the diagnostic clusters are merely results of children who were born in birth clusters moving to areas included in the diagnostic clusters.

Examination of key neighborhood-level resources reveals that the region of the adjusted diagnostic cluster has a high level of diagnostic resources: the number of pediatricians per number of children aged 0 to 4 years is 2.00 (z=23.94, p<0.05) times higher than that of California; the number of advocacy organizations per number of children aged 0 to 4 is 37% higher (z=3.23, p<0.05) than that of the state; the percentage of households earning more than \$100,000 per annum (averaged over ZCTAs) in this region is 26.34%, while it is 22.85% (z=1.6, p=0.05) for California; and regional center spending per number of children aged 0 to 4 is approximately \$700, while it is \$500 for the state. While these results do not provide evidence for a causal association between the presence of these resources and autism diagnoses in all of California, increased resources are associated with the observed diagnostic clusters.

Meanwhile, the adjusted diagnostic clusters have higher prevalence of children with highest functioning autism than California as a whole. The percentage of children with autism who are highest functioning in the shaded area of Figure 4 is 23%, while it is 18% for all of California (z=1.79, p<0.05). In contrast, 46% of the cases in the shaded area are categorized as mid functioning compared to 52% in the whole state (z=1.78, p<0.05). The multivariate scan indicates that not only are higher functioning children with autism more prevalent in this area, but there are small, significant clusters of high functioning children with autism (risks around 6 to 8, relative to the risk of children who are high functioning being diagnosed with autism anywhere in California) nested within the adjusted diagnostic clusters

for 2001 onwards (Figure 4.A). In contrast, the relative risk of a child with autism in the lower functioning category being diagnosed within these small clusters is less than 2.

The clusters cross several regional center service area boundaries, including the boundaries of Lanterman, North Los Angeles, Westside and South Central Los Angeles. Thus, diagnostic clusters of children with high functioning autism are not solely artifacts of differential evaluation practices by regional centers.

Finally, Table 2 displays the results of log-linear models that examine the effects of moving to an area with either fewer or more resources on autism diagnosis. It reveals that moving to an area that is 7 or more levels above one's previous residence in terms of pediatrician density, advocacy organization density and greater median income significantly increases one's odds ratio of a subsequent autism diagnosis relative to children whose resource categories did not change. For example, relative to children whose resource categories do not change, children who move from a region with no advocacy organizations (category 1) to a region with many (category 8) have a 42% higher odds of being diagnosed with autism. Comparable movements across categories of SES and pediatrician density also yield significantly higher odds of 45% and 32% respectively. Regional center spending does not have a significant effect. In contrast, moving to an area with fewer resources reduces one's odds ratio of receiving an autism diagnosis, but the effect is not significant. And as expected, there is no relationship between the density of random locations, mobility and the odds ratio of an autism diagnosis, ruling out the possibility that our findings on the effects of neighborhood resources are artifacts of log linear models of density categories.

DISCUSSION

Our results illustrate that there have been clusters of births of children later diagnosed with autism and clusters of autism diagnoses in California, both robust to key individual-level risk factors. In addition, a number of our findings indicate the important role of resources in shaping autism incidence, as the diagnostic clusters do not completely overlap with the birth clusters and are associated with key neighborhood-level resources. The diagnostic clusters are also disproportionately comprised of children with autism who are higher functioning, implicating the possibility of diagnostic expansion moderated through neighborhood-level resources. The fact that significant diagnostic clusters of children with autism in the highest functioning category were found from 2001 onwards is consistent with the argument that diagnostic expansion played a greater role in the later part of our study period (Dakhlallah & Bearman, 2012). Although the association between neighborhood resources and autism prevalence has been noted previously, (King & Bearman, 2011), the evidence in this article further suggests that resources may play a *causal* role in shaping autism incidence. Most striking is the fact that individuals are more likely to be diagnosed with autism when they move into well-resourced neighborhoods relative to individuals whose neighborhood resources do not change. Since children do not all move into well-resourced neighborhoods from the same places and environmental exposures are hypothesized to be most relevant during early gestation (Arndt, Stodgell, & Rodier, 2005) as well as localized (Mazumdar et al., 2010), the increased odds ratio of diagnosis in the mover group should be specific to the new neighborhood.

The findings reported in this article do not fully reject the possibility that environmental toxicants drive some of the risk of autism. The substantial overlap of the birth and diagnostic clusters along with the high levels of neighborhood diagnostic resources in both do not allow for a complete dissociation between the effects of local toxicants and diagnostic factors. Moreover, since there are a plethora of possible toxicants, it is impossible to falsify all hypotheses that researchers have started to explore (Palmer, Blanchard, Steina, Mandell,

& Miller, 2006; Roberts, English, Grether, Windham, Somberg, & Wolff, 2007; Windham et al., 2006). However, the presence of an environmental component would not challenge our results on the effect of neighborhood-level resources on the likelihood of an autism diagnosis or on the presence of diagnostic expansion against a backdrop of increased diagnoses. Regardless of the role of environmental toxicants, diagnostic resources in a neighborhood have an independent effect on autism incidence. It is important to note that if a toxicant(s) is solely responsible for the clustering of autism at time of birth as well as diagnosis, it would have to disproportionately yield cases of autism that are higher functioning in a small, bounded area. It would also have to be geographically localized to the mapped area of Los Angeles, related to neighborhood-level SES or other resource markers, and have an impact even at ages 3 or 4. While such a confounder would affect our results, its existence seems unlikely. Similarly, different mechanisms specific to the time of birth and then to the time of diagnosis may have operated in the same small area, generating our overlapping clusters. But this too seems unlikely. The presence of diagnostic clusters net of individual-level risk factors could also indicate the presence of the social diffusion of autism awareness among physicians, resulting in localized changes in diagnostic practice. However, although geographic differences in diagnostic practices have been found in relation to other diseases (Yiannakoulias, Hill, & Svenson, 2009), our previous findings on the role of social influence in the autism epidemic indicate that this is not the primary driver of autism diagnoses (Liu, King, and Bearman, 2010). For example, having the nearest elementary school or mall to one's home in common with the nearest child with autism increases a child's likelihood of being diagnosed, while sharing the same nearest pediatrician does not have an effect (Liu & Bearman, Forthcoming).

Another important finding is the lack of an effect of DDS regional center funding on the odds of autism diagnosis. It is possible that funding variations simply reflect a differential availability of services at regional centers or differential operational costs, independent of the services offered. In this regard, it is important to note that the diagnostic clusters are located in high SES, high cost of living ZCTAs. In addition, the clusters we observe cross regional center catchment areas, indicating that the spatial structure of autism in California is not due to variability in diagnostic practices across regional centers.

The spatial scans we used in our analyses have multiple advantages over a stratified spatial analysis. When a geographic area is divided into multiple regions before scanning, each of the scans tends to find the most likely cluster in the most urban area simply because of the increased power offered by urban agglomerations (Waller, Hill, & Rudd, 2006). Therefore, a scan over a large area with multiple urban centers, such as we have done here, is a better methodological alternative, because it implicitly adjusts for urbanicity. It also prevents the arbitrary division of a geographic area into smaller units, which could mask significant effects, making our method conceptually preferable as well.

There are a few limitations associated with our study. Because we use data from California, the generalizability of our results is limited. Also, while the DDS serves the vast majority of children with autism in California, it is not possible to determine whether children with autism who do not utilize DDS services have different patterns of diagnostic risk. In addition, although we adjust for the key known autism risk factors, it is possible that residential sorting according to an unobserved risk factor(s) at the time of birth and at the time of diagnosis contributed to the birth and diagnostic clusters we observe. The key risk factors adjustment in our cluster analyses was done by calculating expected rates of autism using median summarized risks across local areas. While we believe this approach preserves sufficient variability for these analyses, there are other methods, especially in the context of spatial filtering (Rushton & Lolonis, 1996), that allow the adjustment to be done at the

individual level using Monte Carlo Simulations (Banerjee, 2007). Testing spatial patterns using these methods would be interesting avenues of future research.

This study identifies birth and diagnostic clusters of autism in California and implicates a causal relationship between neighborhood-level diagnostic resources and spatial patterns of autism incidence. Consequently, future research that examines contextual factors in relation to autism should be cognizant of distinguishing between clusters at time of birth and at time of diagnosis, as failure to do so could result in spurious conclusions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

REFERENCES

Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. Nat Rev Genet. 2008; 9(5):341–355. [PubMed: 18414403]

Ananthakrishnan AN, Hoffmann RG, Saeian K. Higher Physician Density is Associated with Lower Incidence of Late-stage Colorectal Cancer. Journal of General Internal Medicine. 2010; 25(11): 1164–1171. [PubMed: 20658268]

Arndt TL, Stodgell CJ, Rodier PM. The teratology of autism. International Journal of Developmental Neurosciencer. 2005; 23:189–199.

Baio J. Prevalence of Autism Spectrum Disorders — Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2012. MMWR. 2002; 61(3)

Banerjee A. Temporal changes in the spatial pattern of disease rates incorporating known risk factors. Social Science & Medicine. 2007; 65(1):7–19. [PubMed: 17509742]

Breen. Statistical Models of Social Fluidity. Sociological Theory and Methods. 2006; 21(2):215–236.

Breen. Educational Expansion and Social Mobility in the 20th Century. Social Forces. 2010; 89(2): 365–388.

Breen, Figueroa JB. Stage of breast and cervical cancer diagnosis in disadvantaged neighborhoods: a prevention policy perspective. American Journal of Preventive Medicine. 1996; 12(5):319–326. [PubMed: 8909640]

Bresnahan M, Li G, Susser E. Hidden in plain sight. International Journal of Epidemiology. 2009; 38(5):1172–1174. [PubMed: 19797336]

Charman T, Pickles A, Chandler S, Wing L, Bryson S, Simonoff E, et al. Commentary: Effects of diagnostic thresholds and research vs service and administrative diagnosis on autism prevalence. International Journal of Epidemiology. 2009

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

Dakhlallah D, Bearman P. The expanding definition of severity of autism in California 1993-2002. Manuscript in Progress. 2012

Eng TR, Maxfield A, Patrick K, Deering MJ, Ratzan SC, Gustafson DH. Access to Health Information and Support A Public Highway or a Private Road? JAMA: The Journal of the American Medical Association. 1998; 280(15):1371–1375. [PubMed: 9794322]

ESRI. Community Sourcebook America. Earth Sciences Research Institute; Redlands,CA: 2000-2005.

Eyal, G. The Autism Matrix. Polity; Cambridge: 2010.

Fountain C, Bearman P. Risk as Social Context: Immigration Policy and Autism in California. Sociological Forum. 2011; 26(2):215–240. [PubMed: 21643443]

Kaldor J, Khlat M, Parkin DM, Shiboski S, Steinitz R. Log-Linear Models for Cancer Risk Among Migrants. International Journal of Epidemiology. 1990; 19(2):233–239. [PubMed: 2376429]

Keyes KM, Susser E, Cheslack-Postava K, Fountain C, Liu K, Bearman PS. Cohort Effects Explain the Increase in Autism Diagnosis Among Children Born from 1992 to 2003 in California. International Journal of Epidemiology. 2011

King MD, Bearman P. Diagnostic change and the increased prevalence of autism. International Journal of Epidemiology. 2009; 38(5):1224–1234. [PubMed: 19737791]

King MD, Bearman PS. Socioeconomic status and the increased prevalence of autism in California. American Sociological Review. 2011; 76(2):320–346. [PubMed: 21547238]

King MD, Fountain C, Dakhlallah D, Bearman PS. Estimating autism risk in a time of increasing reproductive age. American Journal of Public Health. 2009; 99(9):1673–1679. [PubMed: 19608957]

Kleinman KP, Abrams AM, Kulldorff M, Platt R. A model-adjusted space–time scan statistic with an application to syndromic surveillance. Epidemiology and Infection. 2004; 133:409–419. [PubMed: 15962547]

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. SaTScanTM v7.0: Software for the spatial and space-time scan statistics. Information Management Services, Inc.; Bethesda: 2006.

Kulldorff, M. SaTScan User Guide v9.1.1. Bethesda,MD: 2011.

Kulldorff M, Mostashari F, Duczmal L, YW K, Kleinman K, Platt R. Multivariate scan statistics for disease surveillance. Statistics in Medicine. 2007; 26(8):1824–1833. [PubMed: 17216592]

Lecavalier, L.; Snow, AV.; Norris, M. Autism Spectrum Disorders and Intellectual Disability. In: Matson, JL.; Sturmey, P., editors. International Handbook of Autism and Pervasive Developmental Disorders. Springer New York; New York, NY: 2011. p. 37-51.

Léonard C, Stordeur S, Roberfroid D. Association between physician density and health care consumption: A systematic review of the evidence. Health Policy. 2009; 91(2):121–134. [PubMed: 19150579]

Leuven, E.; Sianesi, B. Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. 2003.

Liu K, Bearman PS. Focal Points, Endogenous Processes and Exogenous Shocks in the Autism Epidemic. Sociological Methods and Research. (Forthcoming).

Liu K, King M, Bearman PS. Social influence and the autism epidemic. American Journal of Sociology. 2010; 115(5):1387–1434.

Madsen KM, Lauritsen MB, Pedersen CB, Thorsen P, Plesner A-M, Andersen PH, et al. Thimerosal and the Occurrence of Autism: Negative Ecological Evidence From Danish Population-Based Data. Pediatrics. 2003; 112(3):604–606. [PubMed: 12949291]

Marsden PV. Homogenity in confiding relations. Social Networks. 1988; 10:57–76.

Mazumdar S, King M, Liu K, Zerubavel N, Bearman PS. The spatial structure of autism in California, 1993-2001. Health and Place. 2010; 16:539–546. [PubMed: 20097113]

Menken M, Sheps CG. Consequences of an Oversupply of Specialists The Case of Neurology. JAMA: The Journal of the American Medical Association. 1985; 253(13):1926–1928. [PubMed: 3974083]

Mitka M. Rising autism rates still pose a mystery. Journal of the American Medical Association. 2010; 303(7):602. [PubMed: 20159865]

Noterdaeme M, Wriedt E, Höhne C. Asperger's syndrome and high-functioning autism: language, motor and cognitive profiles. European Child & Adolescent Psychiatry. 2010; 19(6):475–481. [PubMed: 19813070]

Oakes, MJ.; Johnson, PJ. Propensity score matching for social epidemiology. In: Oakes, MJ.; Kaufman, JS., editors. Methods in Social Epidemiology. Jossey-Bass; New Jersey: 2006. p. 364-385.

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]

Patterson PH. Immune involvement in schizophrenia and autism: Etiology, pathology and animal models. Behavioural Brain Research. 2009; 204(2):313–321. [PubMed: 19136031]

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. Environmental Health Perspectives. 2007; 115(10)

Roll K. The influence of regional health care structures on delay in diagnosis of rare diseases: The case of Marfan Syndrome. Health Policy (Amsterdam, Netherlands). 2012; 105(2-3):119–127.

Rushton G, Lolonis P. Exploratory Spatial Data Analysis of Birth Defects in an Urban Population. Statisitcs in Medicine. 1996; 15:717–726.

Saulnier CA, Klin A. Social and communication abilities and disabilities in higher functioning individuals with autism and asperger syndrome. Journal of Autism and Developmental Disorders. 2007; 37:788–793. [PubMed: 17160458]

Shipp MPL, Desmond R, Accortt N, Wilson RJ, Fouad M, Eloubeidi MA. Population-based study of the geographic variation in colon cancer incidence in Alabama: relationship to socioeconomic status indicators and physician density. Southern Medical Journal. 2005; 98(11):1076–1082. [PubMed: 16353380]

Vermunt, JK. LEM:A general program for the analysis of categorical data. Department of Methodology and Statistics, Tilburg University; Tilburg: 1997.

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

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(9):1438–1444. [PubMed: 16966102]

Yiannakoulias N. Using population attributable risk to understand geographic disease clusters. Health & Place. 2009; 15(4):1142–1148. [PubMed: 19640774]

Yiannakoulias N, Hill MD, Svenson LW. Geographic hierarchies of diagnostic practice style in cerebrovascular disease. Social Science & Medicine. 2009; 68(11):1985–1992. [PubMed: 19346048]

Zarembo, A. Discovering Autism:Autism boom: an epidemic of disease or of discovery?. Los Angeles Tribune Company; Los Angeles Times: 2011.

Highlights

- In California, there are clusters of births of children later diagnosed with autism and of autism diagnoses.

- The clusters are located in neighborhoods characterized by high socio-economic status and many resources related to child development.

- Children with autism who are higher functioning exhibit stronger clustering than children with autism who are lower functioning.

- Moving to a neighborhood of high socio-economic status or of high resource density increases a child's risk of receiving an autism diagnosis.
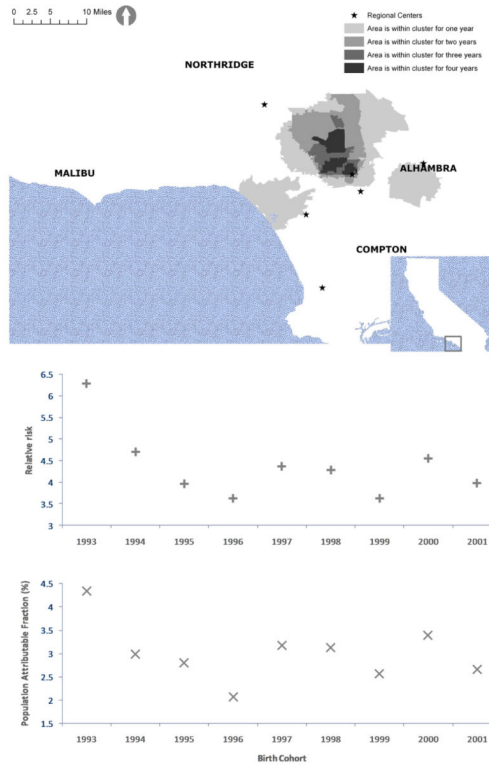
Figure 1.
A) Adjusted autism birth clusters in Los Angeles by temporal duration
B) Risks of a child with autism having been born within a birth cluster relative to all children with autism
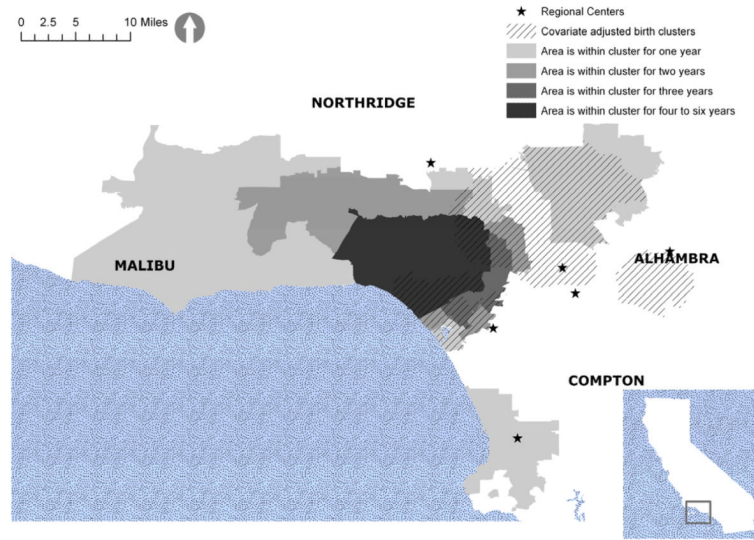C) Population Attributable Fractions of birth clusters

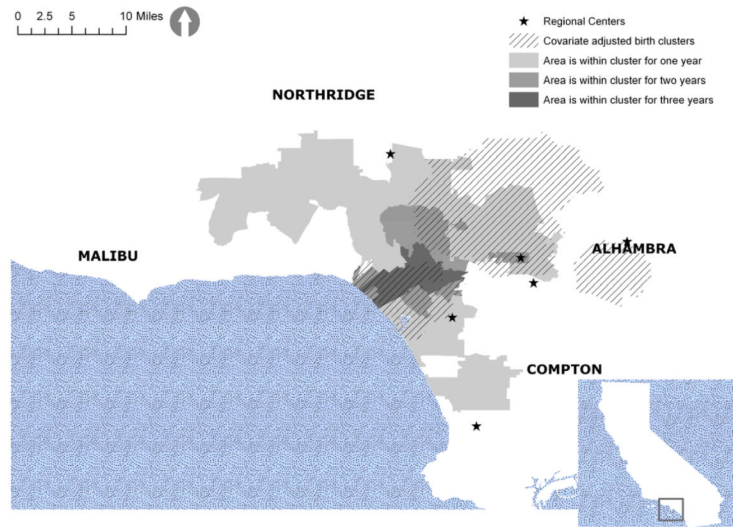**Figure 2.**
Unadjusted autism diagnostic clusters

Figure 3.
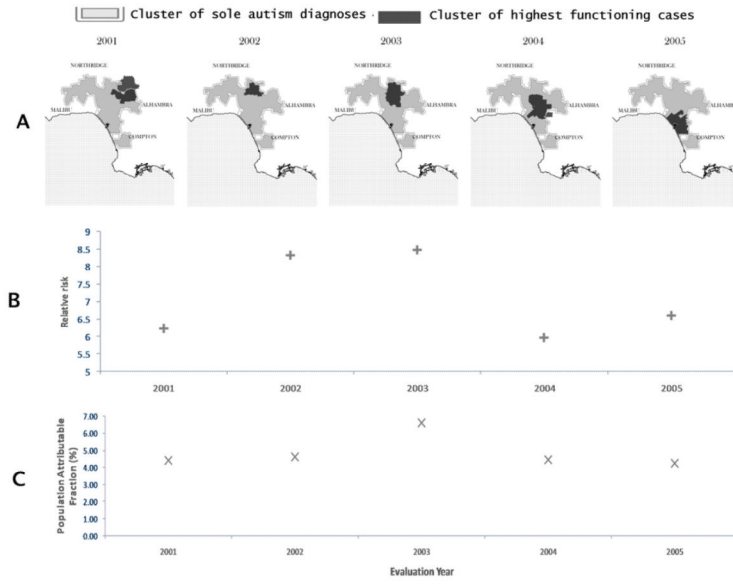Adjusted autism diagnostic clusters

Figure 4.
A) Highest functioning clusters within adjusted diagnostic clusters
B) Risks of a child with autism in the highest functioning category having been diagnosed within a diagnostic cluster relative to all children diagnosed with autism in the highest functioning category
C) Population Attributable Fractions of highest functioning diagnostic clusters

Table 1

Risks of a child with autism having been diagnosed within a diagnostic cluster relative to all children diagnosed with autism and Population Attributable Fractions for diagnostic clusters

Year[a]	Relative risk, unadjusted clusters	Relative risk, adjusted clusters[b]	Population Attributable Fraction, unadjusted clusters	Population Attributable Fraction, adjusted clusters[b]
1998	9.83	7.49	5.32	5.14
1999	4.98	3.71	3.68	2.52
2000	5.03	4.02	3.4	2.55
2001	5.13	4.34	3.62	2.7
2002	4.33	4.56	3.12	3.03
2003	4.37	3.68	3.17	2.59
2004	5.29	4.43	4.07	2.97
2005	4.24	3.81	2.93	2.48

[a] No significant cluster was detected in 1997.

[b] Adjusted for: male sex, Medi-Cal insurance status, preterm birth, low birth weight status, mean parental age, mean parental education, and race/ethnicity

Table 2

Odds ratios of being diagnosed with autism following either downward or upward mobility within 11 categories of pediatrician density, random point density, or advocacy organization density and 10 categories of median income or regional center funding by ZCTA

	Pediatrician density	P-Value	Random point density	P-Value	Advocacy organization density	P-Value	Regional Center funding	P-Value	Median household income	P-Value
Downward mobility	0.99	0.86	0.93	0.34	0.97	0.62	1.04	0.61	0.96	0.52
Upward mobility, Number of categories moved:										
1	1.12	0.09	1.02	0.8	1.20	0.01	1.10	0.13	0.97	0.63
2	0.94	0.48	0.99	0.88	1.17	0.08	0.99	0.88	0.92	0.27
3	0.86	0.10	0.91	0.33	0.99	0.91	0.93	0.48	0.98	0.79
4	1.12	0.24	1.08	0.47	1.15	0.24	0.98	0.89	1.00	0.97
5	1.17	0.15	1.04	0.74	1.06	0.66	0.86	0.45	1.05	0.72
6	1.19	0.16	1.02	0.88	1.27	0.14	0.90	0.71	0.99	0.93
7,8,9,10	1.32	0.03	1.19	0.31	1.42	0.02	1.31	0.46	1.45	0.04
BIC	−636.03		−623.24		−623.24		−485.58		−488.16	