

# Focal Points, Endogenous Processes, and Exogenous Shocks in the Autism Epidemic

Sociological Methods &amp; Research

2015, Vol. 44(2) 272-305

© The Author(s) 2012

Reprints and permission:

[sagepub.com/journalsPermissions.nav](http://sagepub.com/journalsPermissions.nav)

DOI: 10.1177/0049124112460369

[smr.sagepub.com](http://smr.sagepub.com)

Kayuet Liu<sup>1</sup> and Peter S. Bearman<sup>2</sup>

## Abstract

Autism prevalence has increased rapidly in the United States during the past two decades. We have previously shown that the diffusion of information about autism through spatially proximate social relations has contributed significantly to the epidemic. This study expands on this finding by identifying the focal points for interaction that drive the proximity effect on subsequent diagnoses. We then consider how diffusion dynamics through interaction at critical focal points, in tandem with exogenous shocks, could have shaped the spatial dynamics of autism in California. We achieve these goals through an empirically calibrated simulation model of the whole population of 3- to 9-year-olds in California. We show that in the absence of interaction at these foci—principally malls and schools—we would not observe an autism epidemic. We also explore the idea that epigenetic changes affecting one generation in the distal past could shape the precise spatial patterns we observe among the next generation.

## Keywords

autism, simulation, social interactions, environmental risks, diffusion

<sup>1</sup> Department of Sociology, University of California, Los Angeles, CA, USA

<sup>2</sup> Paul F. Lazarsfeld Center for the Social Sciences, Columbia University, New York, NY, USA

## Corresponding Author:

Peter S. Bearman, 701b Knox Hall, 622 W. 122nd Street, Columbia University, New York, NY 10027, USA.

Email: [psb17@columbia.edu](mailto:psb17@columbia.edu)

## **Introduction**

Autism is a serious developmental disorder characterized by impairments in social interaction and communication often accompanied by stereotyped or repetitive behaviors. Its prevalence has increased rapidly in the United States during the past two decades; in California—where reliable data provide a firm foundation for estimation—prevalence increased 631% between the 1992 and 2000 birth cohorts. There is little consensus about the causes of this epidemic. Likewise, we do not understand why the increased incidence of autism is associated with marked—and stable—spatial clustering (Mazumdar et al. 2010). One line of the existing explanations focuses on changes in diagnostic criteria and practice. King and Bearman (2009), for example, show that 24% of the increased prevalence of autism in California can be attributed to the accretion of autism diagnoses that would have previously been diagnoses of mental retardation (MR). Studies looking at sociodemographic factors—most importantly, increased parental age and socioeconomic status—have also identified some broad social trends that have contributed somewhat to the rising prevalence (King et al. 2009; Liu, Zerubavel, and Bearman 2010). Another line of explanation has focused on changes in the environment and/or gene–environment interactions. The effects of environmental toxicants on increased prevalence are far from clear. Those studies suggesting a role for toxicants on autism prevalence have demonstrated modest effects at best (see Windham et al. 2010). Vaccines had been believed to be the major environmental cause of the increase, but the vaccine–autism association has long been refuted by scientific studies (Price et al. 2010; Stehr-Green et al. 2003), and the spatial patterns associated with the epidemic are not consistent with a vaccine link. While millions of dollars have been devoted to identifying a genetic cause of autism, genetics research has yet to identify a genetic factor that can account for more than a small proportion of cases (Abrahams and Geschwind 2008), and the increase in prevalence has occurred far too rapidly for Mendelian dynamics to be thought a salient factor (Liu, Zerubavel et al. 2010). Explanations of the rising prevalence based on gene–environment interaction remain elusive and largely untested.

In sum, the prevailing explanations are unable to account for most of the increase in autism prevalence. While widely acknowledged as a salient factor in the autism epidemic, the impact of increased awareness and knowledge about autism has not often been the focus of empirical studies. We have previously shown that the diffusion of information about autism through spatially proximate social relations has contributed significantly

to the increased prevalence of autism over the 1992–2000 birth cohorts in California (Liu, King, and Bearman 2010). Children do not “catch” autism from someone else, but we observed a wide range of spatially informed empirical patterns that are consistent with local diffusion of awareness of symptoms and the benefits of treatment within parental networks leading to autism diagnoses for children whose development was perceived to be delayed or disordered. These patterns were inconsistent with competing explanations such as shared environmental toxicants, selection, or a physician-led epidemic.

Our previous work focused on the effect of proximity, which has been shown repeatedly in other contexts to be positively associated with the likelihood of social interactions (Haynes 1974). We observed that the risk of an autism diagnosis in the next year increased significantly if a child lived within 500 m of a child previously diagnosed with autism (Liu, King et al. 2010). We also observed that children who moved away from a child with autism were at no additional risk, while children who moved closer to a child with autism were at increased risk for a subsequent autism diagnosis. This fact, coupled with the fact that children living within 500 m of a child previously diagnosed with autism—but in a different school district—were not at additional risk of autism, suggested that parental interactions within neighborhood contexts played a critical role in the diffusion of autism awareness. Increased awareness has in turn led parent to understand developmental dynamics in the language of autism and hence influence each others’ understanding of the signs of autism which may be relevant for their (or their neighbors’) child.

This article expands on this work in four distinct ways. First, we identify the focal points for interaction that drive the proximity effect on subsequent autism diagnoses that we observed previously. Second, we consider how diffusion dynamics through interaction at critical focal points (Feld 1981) shape the dynamics of the spatial distribution of autism cases in California. Third, we expand our observation window to include the most recent data on autism in California through 2010; and finally we consider how an exogenous environmental disaster could have shaped the spatial pattern of the epidemic we observe. We achieve these goals through the development of an empirically calibrated simulation model that allows us to fully explore the social dynamics underlying the autism epidemic in California between 1992 and 2010.

Specifically, for our simulation model we reconstruct the entire population (roughly 3 million per year) of 3- to 9-year-old children in California between 1992 and 2010. Drawing empirical data from three Federal

censuses (1990, 2000, and 2010), the California Birth Master Files, and linking these data to autism data from the California Department of Developmental Services (hereafter, DDS) and location data from private and public sources, we obtain precise measurement on residential and focal point affiliations of our population in temporal context. In this article, we demonstrate that an empirically calibrated simulation model—which takes into account the effects of shared focal points, residential mobility patterns, diagnostic change dynamics, and shifting social demographic population characteristics at the individual and community levels of observation—captures the main contours, both temporally and spatially, of the autism epidemic. Neither changes in diagnostic practices nor social demographic risk factors can account for the temporal or spatial dynamics observed in the autism epidemic. The critical element of the model, however, is the presence of shared interaction foci. In the absence of interaction at these foci—principally malls and schools—we would not observe an autism epidemic.

### *Spatial and Temporal Dynamics*

This finding indicates that the autism epidemic has likely arisen from endogenous social processes. These processes are associated, in California, with a unique temporal and spatial signature. The most striking feature of the spatial signature is that over our entire observation window there is a single spatial birth cluster which is associated with significantly higher risk of autism for children born there (Mazumdar et al., 2010). This area—a  $20 \times 50$  km region above the Hollywood hills in Los Angeles (LA)—is in turn surrounded by 38 smaller clusters where significant additional risk of autism is observed in at least one and often as many as five birth cohorts.

The presence of such birth clusters in the hills above LA—and the absence of such birth clusters in other areas such as the Bay Area (or Silicon Valley more specifically)—enables us to falsify some theories of cause, since the driver of increased prevalence must both make sense of time and place. The presence of an autism birth cluster centered above the Hollywood hills is not a surprising fact to the people who live there. Informal conversation at malls and around schools is sufficient to ascertain both knowledge of the situation and a local theory of cause. The local theory is important—not because it identifies the mechanisms involved, *which it does not*—but rather because it invites a radically new way of thinking about the environment and gene–environment interactions. In this article, we show that this new way to think about environmental risk is not inconsistent with an endogenously driven social epidemic.

In brief, many residents ascribe the increased incidence of autism in their community to current water and/or soil pollution arising from a massive nuclear meltdown that occurred at the Santa Susanna Field Laboratory site (above the communities involved in the autism epidemic today) in 1959. The history of this meltdown and the subsequent cleanup efforts are widely known.<sup>1</sup> For years, the communities with the highest incidence of autism receive water from the LA municipal water system, so source pollution of the drinking water is not selective for the birth cluster. Likewise, for years immediately after the meltdown and quite often subsequently, the Environmental Protection Agency (EPA) has tested the soils in the communities below Santa Susanna for toxicants that could be associated with developmental disorders such as autism. They have not found any. Residents may feel that the EPA is not reporting honestly or testing efficiently, but this seems to be unlikely. However, it is possible that exogenous events such as that which occurred at Santa Susanna Field Laboratory could be associated with the autism epidemic 30 to 40 years later, in the absence of contemporary environmental degradation.

In fact, one of the puzzles of observing an epidemic largely driven by endogenous social processes is to make sense of what such processes would yield clusters in some areas versus others. The simplest answer is that some time in the distant past, the incidence of autism (either by chance or because of a shared exposure) was greater in one area than another. Small inequalities in incidence at one time could easily cumulate into striking differences in incidence when observed a generation later, after social processes have been at play. Recognizing that the striking and significantly different spatial patterns we observe today could reflect very small differences in starting conditions a generation ago also suggests why scholars have been unable to convincingly identify environmental risk factors in the contemporary cross-section. The environmental causes may be there, but they may not express themselves in the autism rate for years, or even generations.

If there is no *current* pollution in the soils and waters of the communities downstream from the Santa Susanna meltdown, this does not mean that the meltdown (and here we actually want to indicate that the logic of explanation operates on a broad array of exogenous environmental events, not a specific event) is unrelated to the autism epidemic. Many mothers of the current children born in the high-risk birth cluster were also born in that cluster, and many were in utero during the meltdown and in the years immediately thereafter. It is possible that exposures at that time led to *de novo* mutations or epigenetic changes that express themselves as autism in the subsequent generation. A very small increase in initial cases, coupled with an endogenous

social influence process, could easily express itself as a high-risk birth cluster years later.

Earlier we noted that the local theory as to why there is an autism cluster is wrong (there is no contemporary soil or water pollution) but that the theory invites a new way of thinking about the environment. In this article we show—through simulation—that small exogenous environmental shocks yielding epigenetic changes, when coupled with a social influence process could in fact lie behind increased prevalence. This is, of course, a thought experiment. But one of the benefits of simulation, especially empirically calibrated simulation, is that it enables us to consider—and potentially reject—such thought experiments. The fact that we cannot reject the possibility of such environmental shocks right away, coupled with the facts that we have not found a contemporary environmental factor associated with increased prevalence and that the environment in the advanced industrial countries where autism is now prevalent is safer now than it was 40 years ago is, at a minimum, suggestive.

### **Road Map**

As noted earlier, the proximity effect on autism is likely to be generated by social interactions at focal points. We empirically examine the impact of sharing the same nearest focal point with one's nearest neighbor with autism, using the same unique sample on which the proximity effect was first identified (Liu, King et al. 2010). To rule out the possibility that the focal point effects we observe are an artifact of proximity, we compare their effects with a series of control locations. While empirical analysis can help us identify the relevant foci, it can only examine the focal points indirectly through spatial proximity, as we lack the information on the actual focal points of these children. The marginal effects from regression models also by definition cannot capture the total impact of an endogenous process. We use a simulation experiment to overcome these limitations. In the simulation, we assign focal points to children according to their sociodemographic characteristics and spatial relationships with actual focal points. This allows us to move beyond marginal effects and simulate the direct impact of focal point interactions on the total increase in incidence. We then consider the impact of exogenous factors such as changes in diagnostic criteria, initial location of cases, and local environmental drivers in tandem with the diffusion processes of the temporal and spatial dynamics of autism.

## **Data and Method**

Our empirical analyses rest on the same sample on which the proximity effect for autism was first identified (Liu, King et al. 2010). From 1997 onward, California's Birth Master Files contain mother's address at birth. Address at birth does not allow us to infer where a child grew up. However, for children with younger siblings, we can infer their residential location for those who reside at the address reported on the younger full sibling's birth record. If the two addresses are different, that is, the family has moved, a child's location in the intervening years cannot be pinned down. Yet for those families who have not moved between the two births, we have an uninterrupted observation window on the elder children with information on residence. To locate the full siblings of our sample, we make an exact match on parents' dates of birth and first initial of mother's maiden name. This generated 1,284,525 potential sibling groups consisting of 2,830,148 children. Among these sibling groups, 533,244 of the eldest children were born at the same address as their next full sibling. A person-year data set was constructed for all these children with no change in address. Excluding the years the child was not at risk (0–2 years old and over 6 years old), the final data set has 321,869 children or 578,925 person-years from 2000 to 2005. The distances of each child in this sample to his or her 10 nearest neighbors with autism in each year were calculated using ArcGIS 9.2 (ESRI, Redland, CA).

In California, the DDS serves the vast majority of children with autism (Croen et al. 2002), and it is these children whose autism diagnoses created an increased probability of an autism diagnosis for those residing in very close proximity to them through the diffusion of information about autism and autism services. We have data on every child in California diagnosed with autism and provided services through the DDS between 1992 and 2010. Individuals served by the DDS are evaluated on a yearly basis. If they move from one residential location to another over the course of the study period, we capture those movements. Only information on those children with autism under the age of 10 was used, as it is unlikely that interactions with parents of adolescents or adults with autism would have much impact on the diagnosis of our very young sample.

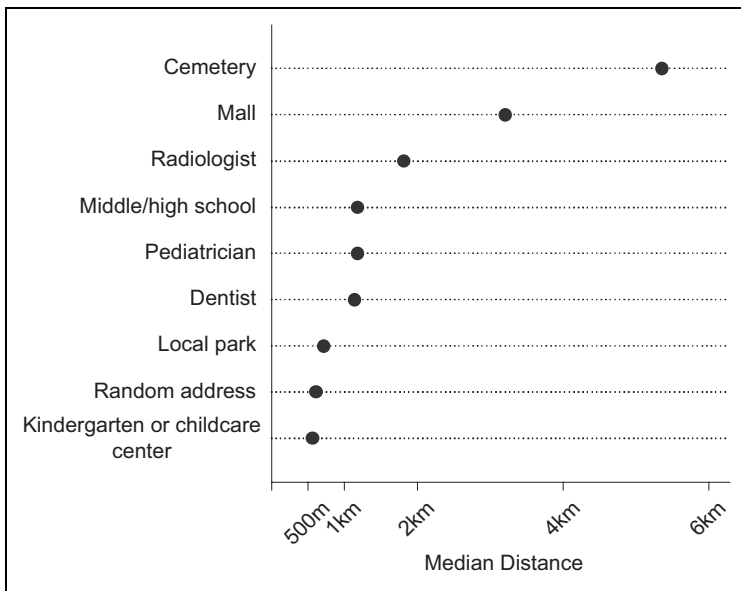
### ***Focal Points***

Parents of small children interact in their neighborhoods at focal points that draw them together. Some of these focal points are obvious: public

elementary schools, licensed child care centers, local parks and major shopping centers (malls). We expected sharing the same nearest focal points with ones' nearest neighbor with autism would lead to an increased chance of a subsequent diagnosis. Because licensed child care centers are often found at the same locations as elementary schools, here we combined the two into one category. Although there is little systematic evidence from which to draw, we do not believe the parent–parent social interactions at pediatrician offices are substantial enough to make them focal points. Rather, being closest to the same pediatrician as ones' nearest neighbor with autism might have an impact on a subsequent diagnosis through a different pathway; the pediatrician of concern may be more likely to diagnose autism or make referrals to developmental specialists. We thus included pediatrician locations to test potential shared diagnostician effects.

All other things being equal, two children closest to the same focal point are more likely to live close to each other than two children who are closest to two different focal points. A focal point effect, therefore, could merely be a spurious effect of proximity. To gauge the extent of confounding, we included some types of places other than focal points as control conditions. We do not expect parents to interact at these control locations. By comparing the performance of the focal points against the control locations, we can rule out the possibility that the focal point effect we observed was merely an artifact of proximity. The control locations we included were randomly selected addresses from the 2000 birth cohort, public middle/high schools, public cemeteries, radiologists and dentists accepting Medi-Cal (California's health welfare system). Random addresses were selected as a control location because they should follow a similar spatial distribution as the sample, and to a lesser extent, the local parks, elementary schools, and child care centers. Middle/high schools can be compared to elementary schools. Cemeteries are not places where parents are likely to talk about child development. The number of public cemeteries is roughly the same as the number of malls; comparing their effects could rule out if the result on malls was due to the smaller number of malls as compared with other locations. Radiologists and dentists were used as control locations for pediatricians; they are places that parents may go, but they are not places where parents will likely talk about their child's development. In fact, people often have trouble talking about anything at the dentist office. We show in the online appendix Table A1 (which can be found at <http://smr.sagepub.com/supplemental/>) the years through which data are available, the number of locations, and the percentage of the sample that had the same nearest location as their nearest neighbor with autism. Figure 1 shows the median distance from the sample





**Figure 1.** Median distance to focal points and control locations in the stayer sample.

to the different types of locations. For each child  $i$ , we created a set of dummy variables to indicate whether he or she had the same closest location of a particular type as his or her nearest neighbor with autism in year  $t - 1$ . We show in the online appendix Table A2 (which can be found at <http://smr.sagepub.com/supplemental/>) the correlation matrix of the dummy variables.

### Statistical Analyses

Discrete event history analysis was used to investigate the effects of sharing a focal point with the nearest neighbor with autism on one's likelihood to be subsequently diagnosed with autism. We estimated the likelihood of being diagnosed with autism during year  $t$  with the following model:

$$\ln\left(\frac{p_{it}}{1 - p_{it}}\right) = \alpha + \sum \beta F_{(t-1)ik} + \sum \theta W_{ik} + \sum \delta Z_{ik}$$

Where  $p_{it}$  equals the probability that child  $i$  will be diagnosed with autism during year  $t$ , given that he or she has not already been diagnosed with autism previously.  $\alpha$ ,  $\beta$ ,  $\theta$ , and  $\delta$  are vectors of logistic regression parameters to be estimated.  $F_{(t-1)ik}$  is  $k$  number of dummy variables indicating whether child  $i$  had the same nearest focal point/control location of a particular type as his or her nearest neighbor with autism in year  $t - 1$ .  $W_{tik}$  and  $Z_{ik}$  are the time-varying and time-invariant control variables, respectively.

Consistent with the literature on autism risk factors, we included the following individual and community-level variables to control for exogenous factors that may affect the probability of an autism diagnosis. Dummy variables for year controlled for the effect of increasing prevalence of autism. Age dummies controlled for the effect of the duration of each child in the person-year data set, which is censored by the birth of the next sibling. Three-year-olds were chosen as the reference category, as three is the most common age at first diagnosis. We also controlled for the key known socio-demographic risk factors for autism—sex, maternal age, and socioeconomic status (Croen et al. 2002; Fountain and Bearman 2011; King and Bearman 2011; Reichenberg et al. 2006). Socioeconomic status was measured by mother's education (in years) and whether the birth and prenatal care was paid for by Medi-Cal. To control for the effect of urbanicity, the density of the population of 0- to 9-year-olds in each school district<sup>2</sup> in year  $t - 1$  (1999–2004) was calculated, based on the 2000 Census and ESRI Sourcebook data (ESRI 2002–2004). The ESRI annual source books contain population estimates for the 0–9 age range projected from the Census data (see ESRI 2007 for details). We interpolated and extrapolated the ESRI and 2000 Census data for the years that data were unavailable (1999 and 2001). The age group of 0–9 population density was logged before entering the model. Logged median income in the school district, also calculated based on the Census and ESRI data, was used to control for the effect of neighborhood-level resources. In short, known risk factors for autism that are likely to be meaningfully spatially distributed and resource variables that could be associated with increased ascertainment are controlled for in our models. We show in the online appendix Table A3 (which can be found at <http://smr.sagepub.com/supplemental/>) the descriptive statistics of the variables.

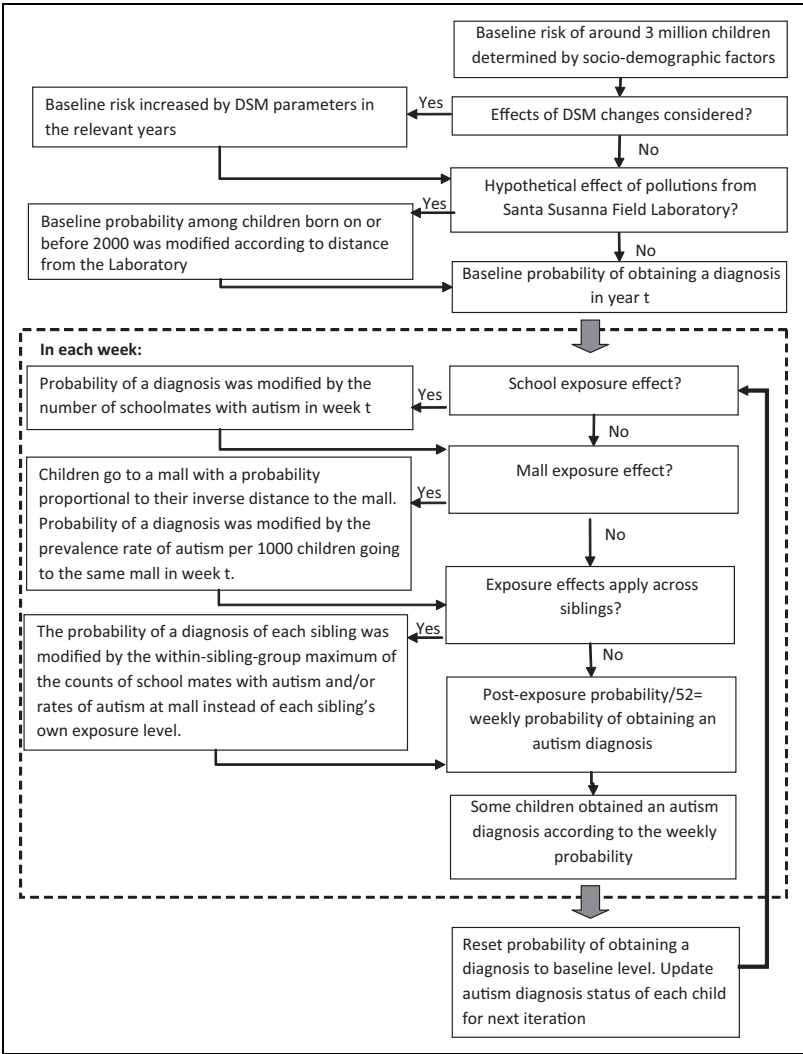
### *Simulation Experiments*

As noted before, the empirical analysis can only examine the effect of sharing the same nearest foci with children with autism, but cannot directly

access the effect of interactions at the focal points, because we do not have information on the actual focal points these children attend. Assigning hundreds of thousands of children randomly to the foci would induce a great amount of error and bias the parameters downward to zero. In a simulation, however, we can assign the children to focal points a priori according to empirical data: sociodemographic characteristics, spatial distances to actual focal points, and if relevant, the carrying capacity (i.e., school capacity) of each focal point. This allows us to simulate the effects of social interactions at foci. In addition, with simulation we can move beyond the marginal effects reported in regression models, which by definition cannot capture the total effect of the endogenous diffusion process that lies behind such effects. For example, while our estimate of the proximity effect indicates that it uniquely accounts for 16% of the observed increase, this is clearly an underestimation of the whole process, which is better conceptualized in dynamic terms captured through simulation. Lastly, we can experiment with different hypothetical scenarios that are particularly relevant to existing theories of the autism epidemic.

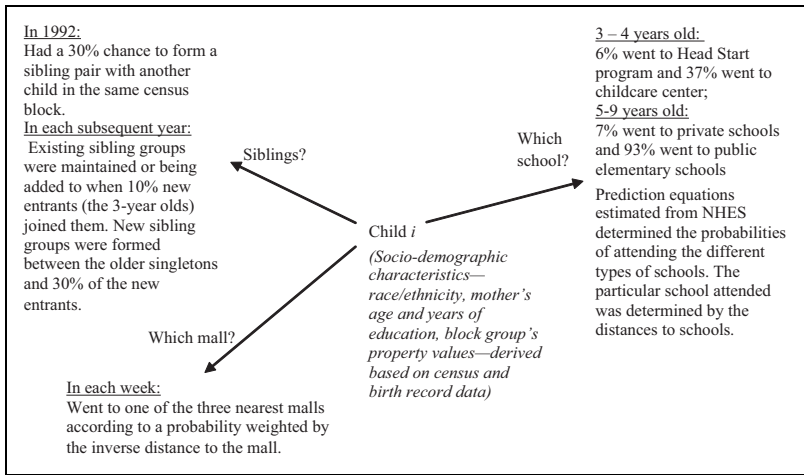
Our simulation model of 57 million person years, located in real space and time, is among the handful of existing large-scale, individual-based microsimulation models (Eubank et al. 2004; Ferguson et al. 2005; Ferguson et al. 2006; Longini et al. 2005) on health outcomes and is the first application of a model of this scale on a noncontagious disease (Galea, Hall, and Kaplan 2009). Infectious diseases can reach pandemic state over a short period of time and, therefore, a static population configuration is usually sufficient. In contrast, noncontagious diseases take a much longer time to diagnose; they also have different mechanisms of diffusion that are less instantaneous. Our model tracks the development of a noncontagious disease over a span of 19 years in a dynamic microsimulation. To maintain a tight coupling between data and models over the long time span, our simulations utilize census and birth record data from multiple years, a residential movement model estimated from 19 years of linked birth record data, and longitudinal enrollment data of schools. To provide an overview, we summarize the endogenous and exogenous processes modeled in the simulation experiments in Figure 2; Figure 3 illustrates how we assigned sibling groups and focal points. The simulations were run on Stata 12 (StataCorp, College Station, TX). The reported simulated incidence rates in this article were the average of 10 simulations of each of the models.

**Population.** We reconstructed the population of all 3- to 9-year-old—about 3 million each year—living in California between 1992 and 2010 based on



**Figure 2.** Exogenous and endogenous processes modeled by the simulation experiments.

block level data from the 1990, 2000, and 2010 censuses. From the three censuses, we have information on the age, race and gender distributions of children living in each census block. We imputed mother's education



**Figure 3.** Assigning children to sibling groups and focal points.

and mother's age based on their distributions by zip code and race groups from the birth record data. Information on property value at the block group level was obtained from the 1990 and 2000 censuses and ESRI updated demographic data for 2010. Population changes were incorporated by gradually replacing the children who aged out of the risk pool with those entering (i.e., reached age 3) in each year. A residential movement model<sup>3</sup> relocated the children throughout the years; We show in the online appendix Table A4 (which can be found at <http://smr.sagepub.com/supplemental/>) the parameter estimates. The model was based on actual movements observed among 2,786,875 children. About half of all children were expected to have moved between age 3 and 9. We show in the online appendix Table A5 (which can be found at <http://smr.sagepub.com/supplemental/>) the descriptive statistics of the synthetic population by year.

Sibling groups were formed to correspond to their empirical occurrence in California over this time period. Specifically, in 1992, pairs were formed among 30% of all children of the same race/ethnicity group and similar mother's age and mother's education level living in the same block. For simplicity, sibling ties were not allowed to form among children of the same age. Then in each subsequent year, 30% of the new entrants to the model (i.e., the 3 years olds) formed ties with the existing children with no ties. Another 10% joined the existing sibling groups following the same race, ethnicity, mothers' age, and education constraints. The resulting proportion

of children with siblings and sibship sizes were comparable to the empirical pattern in the Birth Master Files. Once a sibling group was formed, if any of the members was assigned to move by the residential movement model, their sibling/siblings would also move accordingly. If two members were assigned to move to different zip codes, the assigned movement of one of them was randomly selected as the destination for all of the siblings in the group.

**Baseline Risk.** The baseline risk ( $p_t$ ) of each child to be diagnosed with autism in a particular year ( $t$ ) was determined by the following known risk factors of autism: age, gender, race/ethnicity, mother's age, mother's education, and logged property value. The parameters were derived from a logistic regression model fitted to the 1992 birth cohort and the age of diagnosis information in the DDS data:

$$\begin{aligned} \text{Ln}(p_t/1 - p_t) = & -15.65 + \text{age4} * (-.423) + \text{age5} * (-.79) \\ & + \text{age6} * (-1.44) + \text{age7} * (-2.04) + \text{age8} * (-2.43) \\ & + \text{age9} * (-2.81) + \text{male} * (1.46) + \text{Hispanic white} * (-0.14) \\ & + \text{African American} * (0.54) + \text{mother's age} * (0.04) \\ & + \text{mother's year of education} * (0.09) \\ & + \text{logged property value} * (0.37) \end{aligned}$$

The intercept was set to produce the observed incidence in 1992 and remained constant throughout the simulation. In other words, any increase in incidence over time in the simulation results would be due to compositional changes the social demographic profiles of the population, and/or the social processes tested in each specific model.

### *Endogenous and Exogenous Processes*

**Focal point model.** In the focal point model, the chance of being diagnosed with autism was modified by exposure to other children with autism due to the diffusion of information about autism among parents. Data on actual school locations and capacity were used to assign children to schools. In each year, around 35% of the 3-year-olds and 52% of the 4-year-olds either attend Head Start<sup>4</sup> (6% of all 3- to 4-year-old), or regular licensed child care centers (37% of all 3- to 4-year-old). The probabilities of attending the two types of preschools were estimated based on the National Household Education Surveys' (NHES) 1999 data.<sup>5</sup> The locations and capacities of Head Start programs and licensed child-care centers in 2009 were obtained from the U.S.

Department of Health and Human Services and the California's Department of Social Services. Given their predicted likelihood of attending either type of preschool, the 3- and 4-year-olds were assigned to a specific program/center that was closest to them. When a child's nearest program/center was filled, they were assigned to the next nearest program/center with a vacancy. When a program/center was close to its maximum capacity, preference was given to the children who lived closer to the school over those who lived far away. Once assigned to a preschool, a child would stay there until s/he reached age 5 and moved on to kindergarten in an elementary school. Those children who moved in the residential moving model were reassigned, conditional on a vacancy, to a nearby preschool.

The same procedure was used to assign children to private and public elementary schools. In the simulation, children aged 5 or above attended either private (7%) or public elementary schools (93%). The probabilities of attending private elementary schools were estimated based on the NHES' 1999 data.<sup>4</sup> The information on locations and capacities of private schools was drawn from California's Department of Education's private school registry in 1999. The information on public schools was obtained from the public school registry from 1992 to 2010.

The simulation was set to run for 52 iterations per year.<sup>6</sup> In each iteration, the baseline risk was modified by the impact of contact with the children with autism in one's preschool/elementary school through two parameters,  $\beta_1$  and  $\beta_2$ :

$$\begin{aligned} \text{Ln}(p_t/1 - p_t) = & \text{baseline risk} + \beta_{1*}(N \text{ of schoolmates diagnosed}) \\ & + \beta_{2*}(N \text{ of schoolmates diagnosed})^2 \\ & + \beta_{3*}(\text{autism rate per 1000 at the assigned mall}) \\ & + \beta_{4*}(\text{autism rate per 1000 at the assigned mall})^2 \end{aligned}$$

The combined effect of  $\beta_{1*}(N \text{ of schoolmates diagnosed}) + \beta_{2*}(N \text{ of schoolmates diagnosed})^2$  was not allowed to be negative. Otherwise it was reset to the baseline risk. Note that, to substantially reduce computational time, the simulation model assumes random mixing within groups and will not be modeling the dyadic relationships between individuals directly. Instead of random mixing, the model parameterization is also consistent with diffusion involving second- or higher degree connections, with teachers being the key nodes is a special case.

Apart from going to school, each child also would come into contact with other children at malls during the week. In California—and elsewhere—

malls often contain play areas set aside for children and have, in many cases, replaced parks as a location for joint unstructured play. Parents at such play areas have often arranged to meet at the mall for play dates, perhaps involving lunch at one of the restaurants providing fast food to visitors. The probability of going to one of three nearest malls was determined by each child's inverse distance to each mall. Given the fact that the number of children that are assigned to go to the same mall is much greater than the average school size, it is more reasonable to model the increase in risk by the prevalence rate of autism at the mall. In each iteration, the autism rate per 1,000 children at each mall was calculated, and the combined effect of contact with a child with autism at a mall was determined by  $\beta_{3*}(\text{autism rate per 1,000 at the assigned mall})$  and  $\beta_{4*}(\text{autism rate per 1,000 at the assigned mall})^2$ . Again the combined effect of these two parameters was not allowed to be negative.

In the full focal point model, any modification on the baseline risk was shared within sibling groups, that is, the baseline risk of each child in sibling group  $s$  is modified by  $\max_s[\beta_{1*}(N \text{ of schoolmates diagnosed}) + \beta_{2*}(N \text{ of schoolmates diagnosed})^2 + \beta_{3*}(\text{autism rate per 1,000 at the assigned mall}) + \beta_{4*}(\text{autism rate per 1,000 at the assigned mall})^2]$ . Substantively, this means that any impact from the knowledge about autism gained through one sibling would be automatically applied to the other sibling/siblings. This modified probability of an autism diagnosis (divided by 52) provided the weekly probability of obtaining an autism diagnosis. Some children would switch over to autism, while others would enter the next week without having transitioned. At the next iteration, the probability was restored back to the pre-interaction, baseline level and subsequently modified again by the updated count of children with autism in the school and at the relevant malls.<sup>7</sup>

**Diagnostic change.** Changes in the diagnostic criteria and diagnostic practices, such as those reflected in the Diagnostic and Statistical Manual of Mental Disorders (DSM) and/or specific California guidelines, were expected to have some impact on the diagnosis of autism. We assigned odds ratios of 1.1 ( $\beta_5$ ) and 1.2 ( $\beta_6$ ) for the periods of 1994–1999 and 2000–2010, respectively, using the period of 1992–1993 as our reference category. These odds ratios summarize the effects observed on autism prevalence in studies that explicitly model the impact of changing diagnostic practice on caseload. The effects on the focal point models of changing these parameters were later explored.

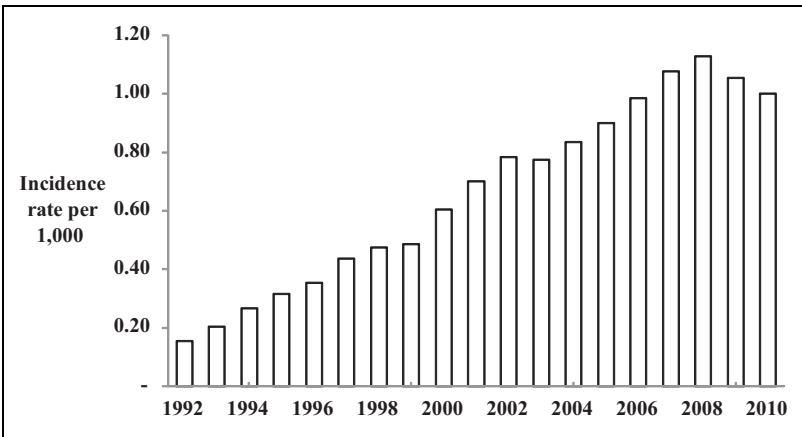
**Seed locations.** Unless otherwise specified, the focal point models were initiated by the actual locations of children with autism in 1992. We explored



the impact of varying the seed locations in the simulation with random locations drawn according to the baseline risk.

*Local theory—pollution from Santa Susana Field Laboratory.* As noted earlier, it is conceivable that some local environmental hazards, even if they happened in the distant past, can affect the spatial distribution of autism through intergenerational genetic and epigenetic processes. Studies of the effects of de novo mutations—genetic mutations on the germ cells—have received much interest in autism research (Lupski 2007; Sebat et al. 2007; Weiss et al. 2008). We have previously proposed that such mutations are one of the genetic mechanisms that can explain the rapid increase in prevalence of a supposedly genetic disorder (Liu, Zerubavel et al. 2010). The time lag between an environmental hazard for autism that acts through de novo mutations and epigenetic pathways would make it difficult to trace the process without substantial—and largely inaccessible—data on the family. However, as a thought experiment, we can assess whether such dynamics could play a role in the autism epidemic through simulation.

We constructed such a scenario by considering the series of environmental disasters occurring at the Santa Susana Field Laboratory in the late 50s through the late 60s. In this hypothetical scenario, some people exposed to the toxicants released from the 1959 Santa Susana Laboratory meltdown (and other accidents) were still living in the nearby areas. We, therefore, modeled the excess risk associated with the toxicant exposures to follow a distance decay function (reported in the online appendix Table A6, which can be found at <http://smr.sagepub.com/supplemental/>) and applied it to all children born in or before 1995—the last birth cohort whose parents may have been exposed to the toxicants released from the laboratory after the series of accidents and to the mishandling of the cleanup process, spanning from late 1950s to late 1960s. Places more than 100 km away from the Santa Susana Field Laboratory, that is, population centers separated by sparsely populated land, were assumed to be unaffected. It is important to recognize that the simulation allows us to assess whether or not an event of this type—not necessarily this specific event—could generate the spatial patterning of the autism that we observe. If so, it may provide rationale for deeper thinking about environmental risk than is reflected in current research, which focuses on exposures contemporaneous with increased prevalence. This is one of the promises of simulation as an analysis strategy; one can falsify ideas by demonstrating through simulation that they would generate outcomes inconsistent with those observed in reality.



**Figure 4.** Incidence rate of autism diagnosis among children aged 3–9 in California, 1992–2010.

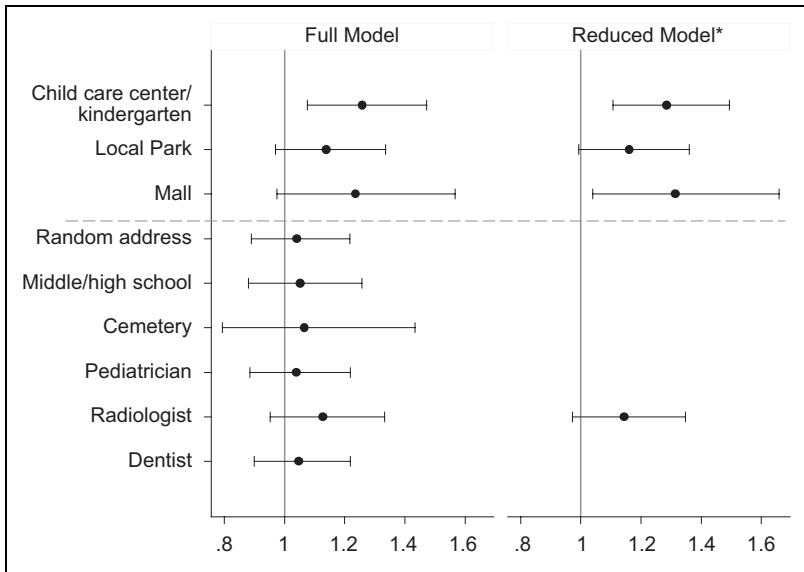
## Results

### *Autism Incidence in 1992–2010*

Figure 4 reports the incidence rate of autism disorder among children aged 3–9 over our observation window based on the DDS data. Information on year and age of first diagnosis was derived from the date of autism diagnosis recorded on the client’s files and did not necessarily coincide with the first year that the client received services from the DDS. This may explain the slight decrease in the incidence rate in 2009 and 2010, as there is, on average, a 1-year gap between the date autism was reportedly diagnosed and the first DDS evaluation in the data. Alternatively, the decline could indicate tightening of diagnostic standards subsequent to California’s fiscal crisis or a decline in the increased caseload.

### *Focal Point Effects in the Stayer Sample*

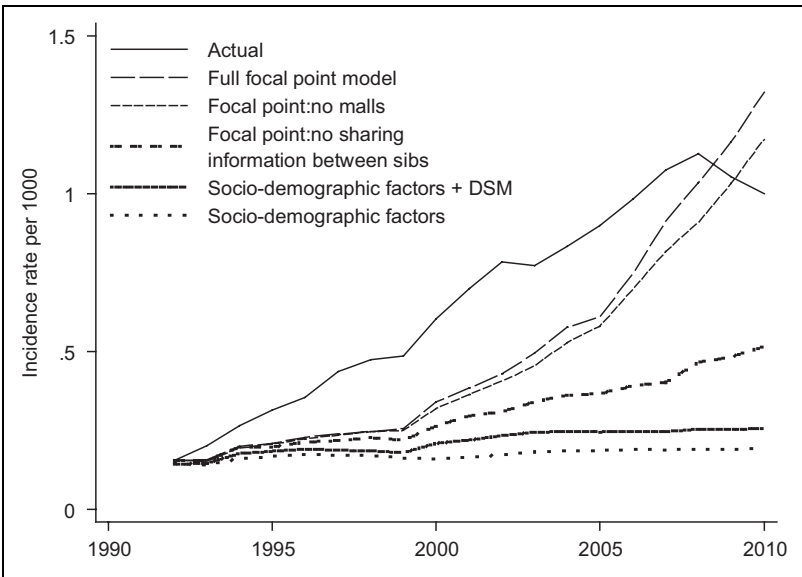
Figure 5 reports the empirical results on focal points, pediatrician, and the control locations (see the online appendix Table A7, which can be found at <http://smr.sagepub.com/supplemental/>, for full sets of parameter estimates). The full model shown in the left panel has all the dummy variables indicating whether a child shared the same nearest focal point as his or her



**Figure 5.** Effects of sharing the same nearest focal points with one's nearest neighbor with autism on a subsequent diagnosis. *Note:* Controlling for year of diagnosis, age sex, race, mother's years of education, mother's age, MediCal, logged population density and logged household income. <sup>a</sup>Includes variable with  $p$  value  $<.2$  in the full model.

nearest neighbor with autism. All the control locations are statistically insignificant. Being closest to the same pediatrician also had no effect. The lack of physician effect on the diffusion of autism is consistent with our previous findings (Liu, King et al. 2010). Notably, being closest to the same elementary schools/child care centers, parks, and malls have strong positive effects. Following parameter reduction convention, the reduced model includes only the variables with a  $p$  value less than .20 in the full model to reduce the level of multicollinearity. The effects of being closest to the same elementary school/child care center and mall are statistically significant in the reduced model.

It should be noted that the correlations between the location dummy variables are only low to moderate and do not have a clear pattern of clustering by whether the locations are focal points or control conditions (see the online appendix Table A2, which can be found at <http://smr.sagepub.com/supplemental/>, for the correlation matrix). Results from principle component



**Figure 6.** Actual and simulated incidence of autism Note: The simulated incidence rates were averages of 10 simulations per model.

analysis (not shown) further support that there is no clear pattern in the associations across the two groups of dummy variables.<sup>8</sup> Hence, the results are not merely artifacts caused by the differential clustering of different types of locations. Similarly, spatial autocorrelation between the variables could have potentially inflated the level of significance. Moran’s *I* statistics were calculated on the Pearson’s residuals of the full model for each of the study years. The *Z* scores range from  $-0.71$  to  $1.5 SD$ , all below the cutoff of a  $\pm 1.65 SD$  at the 95% confidence level. In other words, there is no evidence that there was substantial spatial autocorrelation which might have inflated the significance levels of the parameter estimates.

### Simulation

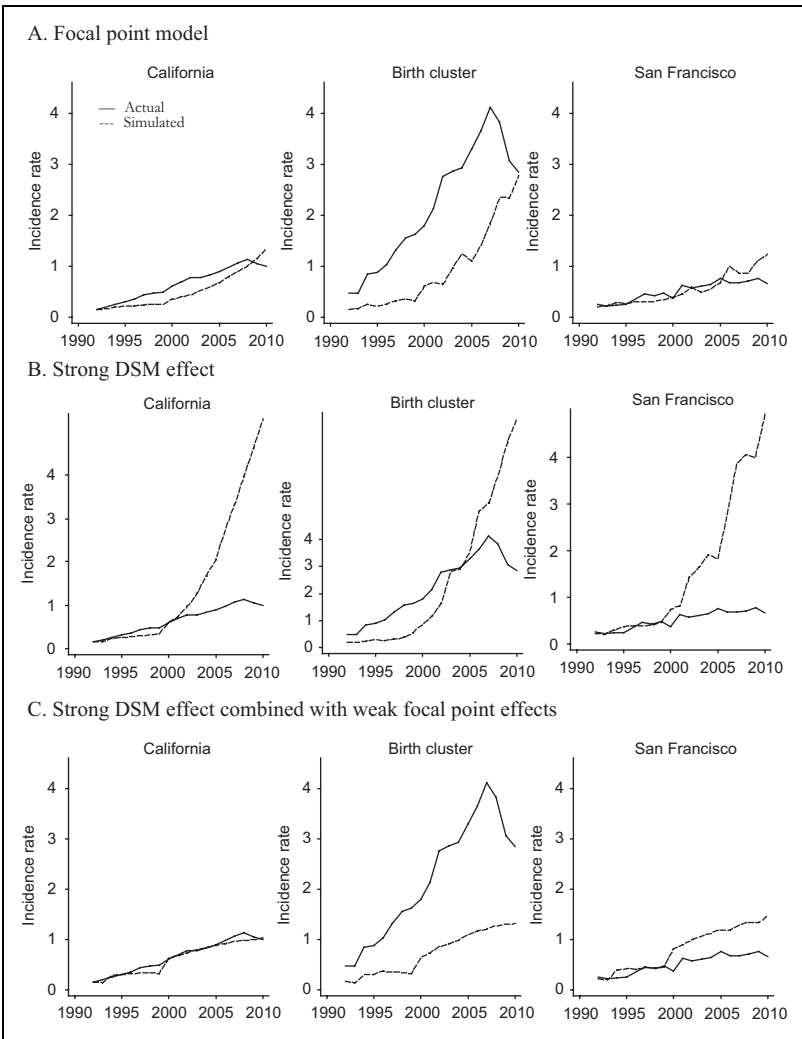
*Effects of focal point interactions versus sociodemographic changes.* Figure 6 shows the actual and the simulated incidence rates between 1992 and 2010. (The parameters of the simulation models shown in Figure 6, as well as all other models reported in this article, can be found in the online appendix

Table A8, which can be found at <http://smr.sagepub.com/supplemental/>). These results clearly indicate that sociodemographic changes in autism risk factors did not generate the epidemic. Note also that changes in diagnostic criteria and practices, at the initial values, also have limited impact. The focal point model, however, captures the contours of the epidemic where we observe rapid increase in incidence. Not surprisingly, the simulated incidence rates are somewhat sensitive to the focal point parameters. We show in the online appendix Table A9 (which can be found at <http://smr.sagepub.com/supplemental/>) the results of varying the school and mall parameters on the simulated incidence.

*Diagnostic Change in Criteria and Practice Corresponds Poorly to the Observed Spatial Pattern.* The top panel of Figure 7 reports the results from the focal point model in more detail. To explore regional differences, Figure 7 also shows the incidence rate in the autism birth cluster identified by Mazumdar et al. (2011) and in an area in San Francisco of a comparable population size. We chose San Francisco because explorative analysis shows that it has an exceptionally low incidence rate, given the average level of baseline risk.<sup>9</sup> Comparing the predicted incidence rates in the two areas, therefore, can test how a model performs in terms of accounting for the spatial dynamics.

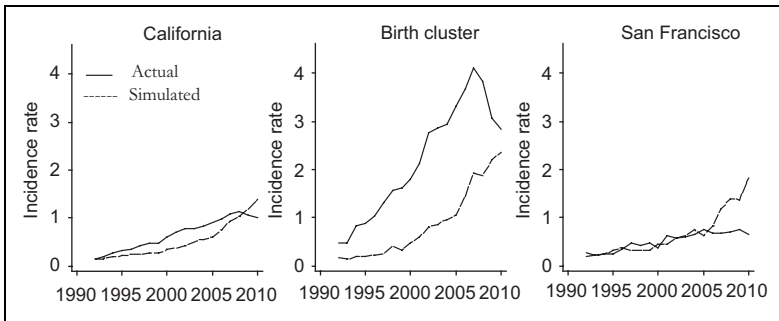
As shown in both Figure 6 and Figure 7, the focal point model underpredicted autism incidence in the earlier years. One might argue that a bigger impact arising from changes in diagnostic practice and criteria could explain the gap between the simulated and actual incidence in the earlier years of the epidemic. Yet as the middle panel demonstrates, any universal increase in risk would be compounded by the diffusion process we identify and lead to exponential increase in incidence. Certainly, the increase in incidence can be brought down by reducing the diffusion parameters. For example, Panel C shows that a strong DSM effect combined with a weak diffusion effect can seemingly reproduce the incidence curve. Yet a closer look at the regional differences shows that such a model fits the regional incidence rates poorly. In other words, while the focal point model can explain, at least in part, the stark regional difference in autism rates, models of universal changes such as diagnostic changes would fail to account for the spatial clustering.

*Effects of Seeding Locations.* It is conceivable that the regional differences in autism incidence in California were the result of an accident—that is the initial distribution of the cases—which, when combined with the social diffusion process, resulted in the present-day spatial pattern. It is also possible



**Figure 7.** Effects of diagnostic changes in interaction with diffusion process.

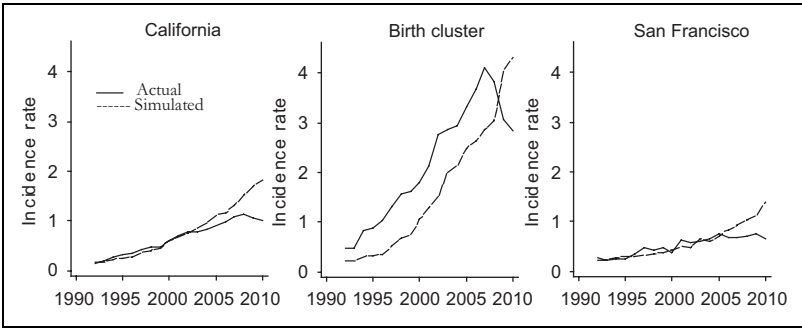
that, had the early autism cases not been where they were, the epidemic could not have taken place. As mentioned above, the focal point model was seeded with the actual locations of the 766 autism cases in 1992. It is meaningful to see whether the temporal and spatial pattern generated by the focal point



**Figure 8.** Random initial locations.

process would change substantially if we changed the locations of the initial cases. In order to assess this, we examined the impact of the initial seed locations by starting the same focal point model with the same number of seeds randomly selected according to the baseline risk in 1992. Note that only the locations of the initial cases were changed; the other aspects of the model, including the baseline risk function, remained unchanged from the focal point model reported in the top panel of Figure 7. As shown in Figure 8, the random seeds generate a spatial pattern resembled that of a model seeded with the actual locations of autism cases. In other words, the spatial pattern generated by the actual locations is not a sporadic event but is in line with the spatial distribution of autism risk.

*Local Theory—Santa Susana Field Laboratory Effects.* The experiment with diagnostic changes suggests that if there were any external factors that account for the discrepancy between the empirical data and the focal point model, they are more likely to be local factors than a universal treatment. The experiment with the seeding schemes suggests that the spatial pattern we observe is not the consequence of merely sporadic events. Here we consider whether our model fits are improved—both in terms of count and spatial distribution—by introducing a local treatment that could contribute to regional differences in case load. Building on the additional risk parameters discussed previously, here we consider whether early (in utero or during the first few years of life) exposure to environmental toxicants released from the Santa Susana Field Laboratory from the late 1950s to the late 1960s increased the risk of giving birth to a child with autism in or before the year 1995.



**Figure 9.** Hypothetical scenario: pollution from Santa Susana Field Laboratory.

Figure 9 shows that the inclusion of this local driver in the simulation leads to a much better fit overall with respect to incidence and with respect to the birth cluster previously identified as an area of heightened risk. Note that the exposure factors have a limited duration in this model and that the social interaction effects—captured through the focal points model structure—sustain the heightened autism rate in the birth cluster even after excess risk is assumed to have dissipated. Certainly, we have no means to determine whether the toxicants released from Santa Susana Field Laboratory have or have not actually contributed to the autism epidemic, or whether there exist some unmeasured, contemporary environmental factors underlying the high incidence at the birth cluster. Yet the thought experiment demonstrates that a temporally distant environment hazard is *capable* of generating marked spatial patterns that are then sustained over time through local influence dynamics, here captured through the focal point model.

*Comparing the Simulated Results With the Stayer Sample.* We compared the simulated results from the focal point model with the stayer sample as an additional robustness check. It seems at first glance that if the focal point model represented the drivers of the autism epidemic in California during the period 1992–2010, the simulated pattern we observe should be comparable to what we observed in the stayer sample. Yet comparing the simulated data and the stayer sample is not straightforward. First, a limitation of the existing empirical data is that the temporal resolution for time of diagnosis is less granular than we would want. The best resolution that could be obtained was year of diagnosis. In contrast, the simulations were structured



around weeks instead of years to allow the endogenous processes to fully play out. Second, as mentioned above, without information on the schools and other focal points the children attended, we cannot fit a focal point model using the stayer sample that is identical to the one we used in the simulation model.

Nonetheless, we can identify a middle ground around which comparable models can be fitted to the empirical and simulated data. The first step would be to approximate the school exposure effect in the simulation model in the stayer sample. We achieve this by calculating the number of times a child was located inside a child with autism's "influence circle" in year  $t - 1$ . An influence circle is defined as an area in which 500 children aged 3–9 resided, which was around the average school size in the simulation model and corresponds to the average elementary school size in California. We thus create a comparable "catchment area." We then regressed the probability of an autism diagnosis in year  $t$  on this measure and its square term using the stayer sample. Second, we calculated the number of schoolmates diagnosed with autism in year  $t - 1$  in the focal point simulation and used it and its square term as predictors of autism diagnosis in year  $t$ . The result is that the two types of models, (A) and (B) in Table 1, are comparable substantively. The effects of sociodemographic factors were controlled for in all models. Table 1 reports that marginal effects of changes in exposure in the stayer sample, the focal point model, and the Santa Susana model. It shows that the focal point models are highly consistent with the empirical data, providing more confidence that we have identified the central dynamic at play.

## **Discussion**

No one has much of a clue as to why autism prevalence has increased so precipitously over the past three decades. The central idea in this article is that one of the reasons that this increase has been so puzzling is that people have been looking at the wrong suspects. Most obviously, by failing to identify the endogenous process by which new cases are identified through the diffusion of increased awareness of both developmental dynamics and treatment opportunities, researchers are stuck trying to find the cause of increased prevalence either in the changing expression of known risk factors or in changing risk factors. Because it seems improbable that epigenetic factors arising from toxicant exposures in previous generations could cause the size of the increases in caseload we observe today, researchers tend to look for changed risk factors in the contemporary period. But by observing that a cascade of

**Table 1.** Changes in the probability of receiving an autism diagnosis in year  $t$  associated with a one unit-change in: (A) the number of times being inside a child with autism’s influence circle in year  $t-1$  and (B) the number of schoolmates diagnosed with autism in year  $t-1$  in the simulation.

|                                | Change in the number of children with autism in year $t-1$ | dy/dx  | 95% C.I. |        |
|--------------------------------|--|--------|----------|--------|
|                                |  |        | lower    | upper  |
| (A) Stayers                    | 0 to 1   | 0.0003 | 0.0002   | 0.0004 |
|                                | 1 to 2   | 0.0004 | 0.0002   | 0.0006 |
|                                | 2 to 3   | 0.0005 | 0.0002   | 0.0008 |
|                                | 3 to 4   | 0.0006 | 0.0001   | 0.0011 |
|                                | 4 to 5   | 0.0008 | 0.0001   | 0.0014 |
|                                | 5 to 6   | 0.0010 | 0.0000   | 0.0019 |
|                                | 6 to 7   | 0.0012 | -0.0001  | 0.0026 |
|                                | 7 to 8   | 0.0016 | -0.0003  | 0.0034 |
|                                | 8 to 9   | 0.0020 | -0.0006  | 0.0046 |
|                                | 9 to 10  | 0.0025 | -0.0010  | 0.0061 |
| (B) Focal point                | 0 to 1   | 0.0002 | 0.0002   | 0.0002 |
|                                | 1 to 2   | 0.0003 | 0.0003   | 0.0003 |
|                                | 2 to 3   | 0.0004 | 0.0004   | 0.0004 |
|                                | 3 to 4   | 0.0005 | 0.0005   | 0.0005 |
|                                | 4 to 5   | 0.0007 | 0.0007   | 0.0008 |
|                                | 5 to 6   | 0.0010 | 0.0009   | 0.0010 |
|                                | 6 to 7   | 0.0013 | 0.0012   | 0.0013 |
|                                | 7 to 8   | 0.0016 | 0.0015   | 0.0017 |
|                                | 8 to 9   | 0.0021 | 0.0019   | 0.0022 |
|                                | 9 to 10  | 0.0025 | 0.0024   | 0.0027 |
| (B) Focal point + Santa Susana | 0 to 1   | 0.0003 | 0.0003   | 0.0003 |
|                                | 1 to 2   | 0.0004 | 0.0004   | 0.0004 |
|                                | 2 to 3   | 0.0005 | 0.0005   | 0.0005 |
|                                | 3 to 4   | 0.0007 | 0.0007   | 0.0007 |
|                                | 4 to 5   | 0.0009 | 0.0008   | 0.0009 |
|                                | 5 to 6   | 0.0011 | 0.0011   | 0.0012 |
|                                | 6 to 7   | 0.0014 | 0.0013   | 0.0014 |
|                                | 7 to 8   | 0.0017 | 0.0016   | 0.0018 |
|                                | 8 to 9   | 0.0020 | 0.0019   | 0.0021 |
|                                | 9 to 10  | 0.0023 | 0.0022   | 0.0024 |

cases could arise from an endogenous influence process operating through sharing information among parents at focal points where parents are likely to discuss their children’s development, the “changed risk factor” problem

becomes less difficult. All one really needs is to identify a risk factor that could have generated a small number of cases to yield an epidemic. But these small numbers of cases cannot just arise anywhere. If they were to have been seeded in the wrong places, one might not generate an epidemic, or the epidemic that one generated would not have the spatial contour of the epidemic we observed.

Simulation studies allow one to ask what could be, i.e., to observe conditions that are thinkable but may not be easily observed in real life. These studies have value in so far as the basic elements of the simulation model are tightly calibrated to empirical data. Such calibration has been the central focus of the research effort that provides the scaffold for the results presented in this article. Critical for our purposes, of course, is the careful work that has been done in the autism research field toward identifying known risk factors for autism. Likewise, central to this effort has been the long research tradition that has focused on the role of changes in diagnostic practice and criteria in the ascertainment of autism and other developmental disorders. These factors—those that shape risk at the individual and community level as well as those that shape the diagnostic regimes that govern the identification and classification process—turn out to be important for individuals, but they are not important for structuring the shape and contour of the autism epidemic in California. No one doubts that increased parental age is a risk factor for autism. Likewise, there is no question that changing diagnostic practices have led to the accretion of autism diagnoses from those with MR diagnoses on the most severely impacted tail of the distribution of autism cases. But these processes—however important they are for shaping individual risk—are not driving the precipitous increase in caseload. Something else is.

Here we see the shadow of what that “something else is.” We also see what it is not. The shadow we observe is captured by the fact that children whose parents share focal points with parents of a child diagnosed with autism are significantly more likely to be diagnosed with autism in the next year than those who do not share such focal points. More precisely, though, we see that only some focal points matter. Parents may meet at dental offices, congregate at cemeteries, and live near to the same random address, radiologist, and pediatrician. But sharing these locations has no effect on the transitions their children may make with respect to autism. It is not the fact of shared foci that matters, it is what must happen at some of the foci that are shared that matters.

We do not know what parents talk to each other about at the play areas in malls or at the school door when they come to pick up their children. But

these are places where parents meet as parents, and it is reasonable to imagine that they see each others' children, talk about their own children, the things they are doing, and their development overall. And it seems reasonable that awareness of autism passes between parents in these conversations; awareness that it is treatable, awareness that one can get resources to help relieve the cost of special services, and awareness of how to navigate a complex bureaucratic service organization (the DDS) to do what all parents want to do—which is to provide for their children in the best way that they can. The focal point model we report in this article provides us with the shadow of the conversations that lie behind the increase in autism prevalence. This is, of course, testable directly; in this regard simulation can provide direction for qualitative study of the dynamics of information diffusion in the context of autism. Without the simulation results though, one might not know to begin to look, or where to look.

It is customary to read that coincident with an increase in autism prevalence one can observe the increase in some other factor which may be construed, because of the temporal correlation of the two time series, to be associated with autism. One could immediately arrive at the absurd idea that frozen yogurt or emo music is associated with autism. They are not, but it is not too far a step to be as concerned about the rise in high-tension wires, or traffic, or TV watching, or other factors that are temporally correlated with increased autism prevalence. This article provides a reason to reject these kinds of arguments. In this simulation process, we observe that capturing the strikingly local variation in autism prevalence—controlling for the sociodemographic risk factors at both the individual and community level—with a global treatment is not possible. Instead, one has to think about how local variation could come about.

Here we show through a thought experiment that a very short exposure associated with modest levels of increased risk for a limited period of time occurring in the generation prior to the generation most at risk for autism—children born between 1992 and 2005—could generate, with the focal point model, an almost exact match to the spatial contours of the epidemic. That such an exposure occurred is of course tantalizing, but the value of the exercise is not to identify a single source of the epidemic. That is not possible; there is no such source. That said, a generation ago, there may have been many small environmental disasters that led to slightly elevated incidence of developmental disorder. These disasters, local, but scattered here and there in the developed world, could, when coupled with an endogenous dynamic influence process, have led to the autism epidemic we observe today.

Simulation exercises can never prove such an idea. But they can falsify such an idea. That our simulation falsifies the idea that the epidemic is caused by pediatrician offices, changes in diagnostic criteria and/or practice, and conversations that happen where people do not talk about children, but fails to falsify the idea that epigenetic changes induced by short local exposure to a toxicant with developmental implications is worth thinking about.

Of course, this study has many limitations. Our empirical data arise from California; we do not know whether parents actually talk to each other; and our results are built of a simulation of social demographic dynamics resulting in family formation and residential moving. As with all simulations, the results are not real—the results are facts that are good to implement in our thinking, but which we may meaningfully contrast with facts that are facts in and of themselves.

### **Acknowledgments**

Christine Fountain, Alix Winter, Kinga Makovi, Keely Cheslack-Postava and Soumya Mazumdar contributed enormously useful assistance for which we are deeply grateful.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is supported by the National Institutes of Health (NIH) Director's Pioneer Award program, part of the NIH Roadmap for Medical Research (Grant 1 DP1 OD003635-01).

### **Notes**

1. The Santa Susana Field Laboratory was used for the testing and development of rocket engines and nuclear reactors. In 1959, a partial nuclear meltdown releasing roughly 400 times the amount of radiation released at Three Mile Island occurred. The subsequent cleanup of the site involved shooting cans of radioactive materials collected after the meltdown with small arms fire, thereby spreading contamination into the atmosphere. Other nuclear accidents happened in the 1960s at the site and another radioactive fire occurred in 1971.

2. The Los Angeles unified school district was further partitioned into eight subdistricts based on the district's local boundaries.
3. A logistic regression model predicting whether a child  $i$  born between 1985 and 2007 had a different zip code at birth as compared to his or her younger sibling used. Only moves made before child  $i$  reached age 10 were considered; 46.64% of this sample had a cross zip code move before age 10. The origin–destination zip code matrices from four periods (1985–1989, 1990–1994, 1995–1999, and 2000–2007) were calculated. In the simulation, the parameter estimates reported in the online appendix Table A2 (which can be found at <http://smr.sagepub.com/supplemental/>) were used to predict the odds of moving in the simulation and the destination was determined by the four origin–destination zip matrices. Note that the residential movement model was estimated based on only children born in California and moving within the state; potential differences in the residential movement patterns by whether the child was born in California, as well as the movements from and to out of California, have not been considered.
4. A Federal program for preschool children from low-income families.
5. The NHES program is one of the few national surveys that contain education data on very young children. We fit logistic regression models to predict attending license child care, Head Start program and private elementary schools using the 1999 data and yielded the following predictive equations:

$$\begin{aligned} \text{Ln}(p_{\text{child care}}/1 - p_{\text{child care}}) = & -1.64 + 0.7(4 \text{ years old}) \\ & + 0.35(\text{percentage of mothers with children} \\ & \text{aged 5 or below working in census block group}) \\ & - 0.78(\text{Hispanic white}) + 0.56(\text{mother graduated} \\ & \text{from high school}) + 1.02(\text{mother had some college education}) \\ & + 1.37(\text{mother graduated from college}) \\ & + 0.22(\text{had a household income of } 35 - 50\text{k}) \\ & + 0.42(\text{had a household income of } 50 - 75\text{k}) + \\ & 0.87(\text{had a household income } > 75\text{k}) \end{aligned}$$

$$\begin{aligned} \text{Ln}(p_{\text{Head Start}}/1 - p_{\text{Head Start}}) = & -2.57 - 0.8(4 \text{ years old}) \\ & + 1.58(\text{African American}) + 0.86(\text{Hispanic white}) + 0.65(\text{other races}) \\ & - 0.65(\text{mother graduated from college}) - 0.44 (\text{had a household income} \\ & \text{of } 25 - 35\text{k}) - 0.66(\text{had a household income of } 35 - 50\text{k}) \\ & - 0.88(\text{had a household income of } 50 - 75\text{k}) \\ & - 1.54(\text{had a household income } > 75\text{k}); \end{aligned}$$

$$\begin{aligned} \text{Ln}(p_{\text{privateschool}}/1 - p_{\text{privateschool}}) = & -3.38 - 0.11 \\ & (\text{percentage of mothers with children aged 6-17} \\ & \text{working in census block group}) + 0.47(\text{mother graduated from high school}) \\ & + 0.72 (\text{mother had} \\ & \text{some college education}) + 1.23(\text{mother graduated from college}) \\ & + 0.47(\text{had a household income of} \\ & 25 - 35\text{k}) + 0.74(\text{had a household income of } 35 - 50\text{k}) \\ & + 0.76(\text{had a household income of } 50 - 75\text{k}) \\ & + 0.125(\text{had a household income } > 75\text{k}). \end{aligned}$$

6. We chose week as the unit of the iteration to maximize the frequency of updating the information that would have been available to the agents while keeping the computational time of the simulation manageable. For simplicity, we did not consider holidays, which should have limited impact in our case, as children with autism typically do not lose their diagnosis over time. In comparison, holidays have a larger impact in models of infectious diseases for which recovery of the infected plays a role in the diffusion process.
7. By restoring the probability to the baseline level, the focal point model reported here assumed that the effect of social interactions on the chance of a subsequent diagnosis is the same in each iteration. In some other work, we have considered the impact of the duration of social relations on the diffusion of knowledge about autism, but it is beyond the scope of this article.
8. Scree plot of eigenvalues after principle component analysis of the 10 location dummy variables suggests a single component. Inspecting the component loadings of forced 2, 3, or 4 component models shows that the focal points and control locations are equally likely to load on the same components.
9. It is also the case that many genetically based arguments for the rapid increase in autism would identify San Francisco—and the Bay area in general—as likely locations for an autism birth cluster. The general line of argument, summarized as the so-called geek hypothesis, is that individuals on the border of autism spectrum disorder find positions of high prestige in the new technological industrial sectors and through assortative mating dynamics place their children at greater risk for genetic expression of autism. The absence of a cluster in the Bay Area indicates that this theory, while popular in the lay press, is not supported.

## References

- Abrahams, B. S. and D. H. Geschwind. 2008. "Advances in Autism Genetics: On the Threshold of a New Neurobiology." *Nature Reviews Genetics* 9:341-55.

- Croen, L. A., J. K. Grether, J. Hoogstrate, and S. Selvin. 2002. "The Changing Prevalence of Autism in California." *Journal of Autism and Developmental Disorders* 32:207-15.
- ESRI. 2002–2004. *Community Sourcebook—America with ArcReader* (2002-2004 editions). Redlands, CA: ESRI.
- ESRI. 2007. *Evaluating Population Projections—The Importance of Accurate Forecasting. ESRI White Paper* Redlands, CA: ESRI.
- Eubank, S., H. Guclu, V. S. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, et al. 2004. "Modelling Disease Outbreaks in Realistic Urban Social Networks." *Nature* 429:180-84.
- Feld, Scott. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86:1015-35.
- Ferguson, N. M., D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, et al. 2005. "Strategies for Containing an Emerging Influenza Pandemic in South-east Asia." *Nature* 437:209-14.
- Ferguson, N. M., D. A. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, and D. S. Burke. 2006. "Strategies for Mitigating an Influenza Pandemic." *Nature* 442:448-52.
- Fountain, Christine and Peter S. Bearman. 2011. "Risk as Social Context: Immigration Policy and Autism in California." *Sociological Forum* 26:215-240.
- Galea, S., C. Hall, and G. A. Kaplan. 2009. "Social Epidemiology and Complex System Dynamic Modelling as Applied to Health Behaviour and Drug use Research." *International Journal of Drug Policy* 20:209-16.
- Haynes, Robin M. 1974. "Application of Exponential Distance Decay to Human and Animal Activities." *Geografiska Annaler. Series B, Human Geography* 56:90-104.
- King, M. and P. Bearman. 2009. "Diagnostic Change and the Increased Prevalence of Autism." *International Journal of Epidemiology* 38:1224-34.
- King, M. D. and P. S. Bearman. 2011. "Socioeconomic Status and the Increased Prevalence of Autism in California." *American Sociological Review* 76:320-46.
- King, M. D., C. Fountain, D. Dakhllallah, and P. S. Bearman. 2009. "Estimated Autism Risk and Older Reproductive Age." *American Journal of Public Health* 99:1673-79.
- Liu, K., M. King, and P. S. Bearman. 2010. "Social Influence and the Autism Epidemic." *American Journal of Sociology* 115:1387-434.
- Liu, K., N. Zerubavel, and P. Bearman. 2010. "Social Demographic Change and Autism." *Demography* 47:327-43.
- Longini, I. M., Jr., A. Nizam, S. Xu, K. Ungchusak, W. Hanshaworakul, D. A. Cummings, et al. 2005. "Containing Pandemic Influenza at the Source." *Science* 309:1083-87.
- Lupski, J. R. 2007. "Genomic Rearrangements and Sporadic Disease." *Nature Genetics* 39:S43-47.



- Mazumdar, S., M. King, K. Y. Liu, N. Zerubavel, and P. Bearman. 2010. "The Spatial Structure of Autism in California, 1993-2001." *Health and Place* 16:539-46.
- Price, C. S., W. W. Thompson, B. Goodson, E. S. Weintraub, L. A. Croen, V. L. Hinrichsen, et al. 2010. "Prenatal and Infant Exposure to Thimerosal from Vaccines and Immunoglobulins and Risk of Autism." *Pediatrics* 126:656-64.
- Reichenberg, Abraham, Raz Gross, Mark Weiser, Michealine Bresnahan, Jeremy Silverman, Susan Harlap, et al. 2006. "Advancing Paternal Age and Autism." *Archives of General Psychiatry* 63:1026-32.
- Sebat, J., B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, et al. 2007. "Strong Association of De Novo Copy Number Mutations with Autism." *Science* 316:445-49.
- Stehr-Green, P., P. Tull, M. Stellfeld, P. B. Mortenson, and D. Simpson. 2003. "Autism and Thimerosal-Containing Vaccines: Lack of Consistent Evidence for an Association." *American Journal of Preventive Medicine* 25:101-106.
- Weiss, L. A., Y. Shen, J. M. Korn, D. E. Arking, D. T. Miller, R. Fossdal, et al. 2008. "Association Between Microdeletion and Microduplication at 16p11.2 and Autism." *New England Journal of Medicine* 358:667-75.
- Windham, G. C., G. King, E. Roberts, and J. Grether. 2010. "Autism and Distribution of Hazardous Air Pollutants at Birth in Southern California." *International Meeting for Autism Research*, Philadelphia, PA.

## Author Biographies

**Kayuet Liu, DPhil**, is an assistant professor at the Department of Sociology, UCLA. She received her DPhil in sociology from the University of Oxford in 2008 and was a postdoctoral research scholar at the Paul F. Lazarsfeld Center for the Social Sciences, Columbia University. Her research focuses on how large-scale phenomena—from the distribution of suicides at the workplace to the rising prevalence of autism—arise from the interactions among the individuals involved. Mechanisms of information diffusion and the spatial/temporal dimensions of social dynamics are central to her research. She is the first author of the articles "Social Influence and the Autism Epidemic" (2010) published in *American Journal of Sociology*, 115:1387-1434 and "Social demographic change and autism" (2010) published in *Demography* 47(2):327-343.

**Peter Bearman, PhD**, is the director of the Lazarsfeld Center for the Social Sciences, the Cole Professor of Social Science, and codirector of the Health & Society Scholars Program. He was the founding director of ISERP, serving from the Institute's launch in 2000 until 2008. A recipient of the NIH Director's Pioneer Award in 2007, Bearman is currently investigating the social determinants of the autism epidemic and is a coauthor of the article "Social Influence and the Autism Epidemic" (2010) published in *American Journal of Sociology*, 115:1387-1434. A specialist in network analysis,

---

he codesigned the National Longitudinal Study of Adolescent Health and has used the data extensively for research on topics including adolescent sexual networks, networks of disease transmission, and genetic influences on same-sex preference. He has also conducted research in historical sociology, including *Relations into Rhetoric: Local Elite Social Structure in Norfolk, England, 1540-1640* (Rutgers, 1993). He is the author of *Doormen* (University of Chicago Press, 2005).