CrossMark

Neighborhood Disadvantage is Associated with High Cytomegalovirus Seroprevalence in Pregnancy

Paul M. Lantos [1,2,3] · Kate Hoffman [4] · Sallie R. Permar [1,5] · Pearce Jackson [6] ·
Brenna L. Hughes [6] · Amy Kind [7] · Geeta Swamy [6]

Abstract

Background Cytomegalovirus (CMV) is the most common infectious cause of fetal malformations and childhood hearing loss. CMV is more common among socially disadvantaged groups, and geographically clusters in poor communities. The Area Deprivation Index (ADI) is a neighborhood-level index derived from census data that reflects material disadvantage.

Methods We performed a geospatial analysis to determine if ADI predicts the local odds of CMV seropositivity. We analyzed a dataset of 3527 women who had been tested for CMV antibodies during pregnancy. We used generalized additive models to analyze the spatial distribution of CMV seropositivity. Adjusted models included individual-level age and race and neighborhood-level ADI.

Results Our dataset included 1955 CMV seropositive women, 1549 who were seronegative, and 23 with recent CMV infection based on low avidity CMV antibodies. High ADI percentiles, representing greater neighborhood poverty, were significantly associated with the nonwhite race (48 vs. 22, $p < 0.001$) and CMV seropositivity (39 vs. 28, $p < 0.001$). Our unadjusted spatial models identified clustering of high CMV odds in poor, urban neighborhoods and clustering of low CMV odds in more affluent suburbs (local odds ratio 0.41 to 1.90). Adjustment for both individual race and neighborhood ADI largely eliminated this spatial variability. ADI remained a significant predictor of local CMV seroprevalence even after adjusting for individual race.

Conclusions Neighborhood-level poverty as measured by the ADI is a race-independent predictor of local CMV seroprevalence among pregnant women.

Keywords Cytomegalovirus · Health disparities · Poverty · Geographic information system · Spatial epidemiology · Generalized additive model · Pregnancy

✉ Paul M. Lantos
Paul.lantos@duke.edu

1 Department of Medicine, Duke University, DUMC 100800, Durham, NC 27710, USA

2 Department of Pediatrics, Duke University, Durham, NC, USA

3 Global Health Institute, Duke University, Durham, NC, USA

4 Nicholas School of the Environment, Duke University, Durham, NC, USA

5 Human Vaccine Institute, Duke University, Durham, NC, USA

6 Department of Obstetrics and Gynecology, Duke University, Durham, NC, USA

7 Department of Medicine, University of Wisconsin, Madison, WI, USA

Introduction

Congenital cytomegalovirus infection (CMV) is the leading infectious cause of neurologic deficits and hearing loss in infants, resulting in more long-term pediatric disabilities than trisomy 21 and spina bifida [1]. CMV is more common among socially disadvantaged groups and nonwhite minorities [2–12]. We have recently conducted geospatial analyses demonstrating that CMV seropositivity, including among pregnant women, significantly clusters in poor urban neighborhoods with large minority populations in North Carolina [11, 12]. Adjustment for race did not completely abrogate this clustering, indicating that

Springer

geographic concentrations of CMV may be additional social determinants of CMV risk.

To study social determinants of individual health poses numerous challenges, particularly in studies of electronic medical records data, which typically contain little demographic or socioeconomic data. Neighborhood-level socioeconomic variables, however, can be incorporated into models of individual health outcomes when the individuals' geography is known. The Area Deprivation Index (ADI) is a weighted index comprised of 17 census-based markers of material deprivation and poverty [13]. In this study, we have investigated whether the ADI, calculated at the level of census block groups (i.e., "neighborhood"), is associated with individual risk of CMV seropositivity during pregnancy, and whether this is independent of individual race.

Methods

Design and Cohort

This was a cross-sectional case-control study using electronic health records and maternal CMV serologic data. The subjects included in this study are from a previously reported cohort of 3527 women who had been tested for CMV antibodies during pregnancy [12]. These women were Duke University Health System patients who had been screened for participation in a multicenter trial of CMV hyperimmune globulin in pregnant women with recently acquired CMV infection (NCT 1376778). Our dataset combined CMV testing results from this trial with an electronic query of the trial subjects' electronic health records to obtain their age, race, ethnicity, and the coordinates of their home address.

Geographic Data Management

We used the geographic information system (GIS) software ArcGIS 10.3.1 (ESRI, Redlands, CA) for spatial data management and map production. Our initial dataset of 6396 patients included many spatial outliers. We used GIS operations to select a subset with high spatial density, retaining patients whose residential coordinates fell within Durham County, NC, or one of the five bordering counties (Wake, Person, Chatham, Orange, and Granville). Within these six counties, there remained some peripheral areas with very few subjects; thus, to further maximize spatial sampling density, we calculated a two-standard deviation ellipse and retained the subjects it contained. This is the smallest ellipse that contains 95% of patient addresses. Ultimately, 3527 subjects remained in this elliptical study area encompassing parts of six counties.

For the statistical modeling described below, we used 3504 women who could be categorized as either CMV seronegative or as seropositive with high avidity CMV antibodies. The remaining 23 women had low avidity CMV antibodies, indicating recent primary CMV infection. As these women had been seronegative until recently, they were both immunologically and possibly epidemiologically distinct from the remaining CMV seropositive women. Consequently, we did not assign them to either category in our spatial modeling.

Area Deprivation Index

We obtained neighborhood-level ADI scores for all North Carolina census block groups. ADI scores were calculated using the 2013 American Community Survey 5-year averages. Higher ADI values (and therefore percentiles) represent greater degrees of socioeconomic contextual disadvantage [14]. We found that the distribution of ADI values was similar for our study region as for the entirety of North Carolina (Supplement 1). We converted neighborhood ADI values to percentiles using the statewide ADI distribution.

Statistical Analyses

Statistical analyses were performed using the statistical programming language R 3.2.3 (www.r-project.org). We compared mean ADI percentile by (1) CMV serostatus and (2) race using Mann-Whitney tests.

Our primary spatial analysis was a generalized additive model (GAM), which we used to predict a continuous odds ratio (OR) surface over the geographic study area. This was accomplished using the mgcv package in R [15]. The GAM is similar to logistic regression, but with fewer assumptions about the functional form of the relationship. The GAM uses nonparametric splines as a smoothing function to compute local variability in the relationship between our outcome variable and spatial coordinate pairs (as defined by the longitude and latitude coordinates for each study subject). Log-odds are predicted over a dense longitude-latitude grid covering the geographic extent of the study area, then divided by the global odds from a non-spatial model to calculate a pointwise odds ratio. Permutation tests are then used to determine the statistical significance of spatial variation across the study area as a whole and of OR predictions; a two-tailed p of < 0.05 after 1000 permutations is accepted as statistically significant. We constructed two spatial GAM models: an unadjusted model with only our outcome variable and the smoothed spatial parameters, and a full model that also included both individual predictors and ADI percentile. In our prior study using this dataset, we found that the prevalence of CMV seropositivity increased with age among minority women, but remained constant among non-Hispanic white women [12]. In the statistical models, this difference yielded a significant interaction term between patient age and patient race. Consequently, we included an age-race interaction term in the models in this study. The code for our models can be found in Supplement 2.

Results

Patient Cohort

Our study cohort included 3527 women, of whom 1955 were CMV seropositive and 1549 were seronegative (55.7% seropositive, 95% CI 54.1–57.4). In the cohort, 93.4% were either white (1928, 54.2%) or African American (1394, 39.2%). The remainder were Asian (191, 5.3%), Native American (7, 0.2%), and Hawaiian/Pacific Islander (7, 0.2%). Forty-six of these subjects (1.3%) identified as Hispanic, 41 of whom designated their race as "white." We dichotomized our overall cohort into 1887 "Non-Hispanic White" and 1640 "Minority" categories. CMV seropositivity was substantially higher among minorities than among non-Hispanic whites (71.7 vs. 41.9% OR 3.76, 95% CI 3.25–4.34).

ADI

Minority women had substantially higher mean ADI percentile than non-Hispanic white women (48 vs. 22, $p < 0.001$). Overall, the mean ADI percentile was higher among CMV seropositive women (39 vs. 28, $p < 0.001$). Of 23 women with low avidity CMV antibodies, indicating recent infection, the mean ADI percentile was 32.8. The relationship between ADI percentile and CMV serostatus remained statistically significant when each racial group was analyzed independently (for non-Hispanic white women 23 CMV+ vs. 21 CMV−, $p = 0.017$; for minority women 49 CMV+ vs. 46 CMV−, $p = 0.048$).

Spatial Models

Our unadjusted model showed a statistically significant spatial effect compared (global p value < 0.001 compared with a nonspatial model). This revealed marked local heterogeneity of CMV seropositivity, with a cluster of high odds in the urban centers of Durham and Raleigh and clusters of low odds in the surrounding suburban communities (Fig. 1). The local odds ratio of CMV varied from 0.41 to 1.90. Our fully populated model, which included both ADI percentile and the race-age interaction terms, significantly abrogated the spatial heterogeneity and blunted the local OR range to between 0.76 and 1.21. After adjustment for ADI, race, and age, the spatial model was not significantly better than a non-spatial model (global p value = 0.26). ADI percentile and individual race both remained statistically significant in models that included both.

Discussion

We have found that the likelihood of CMV seropositivity among pregnant women is significantly associated with ADI, a neighborhood-level measure of socioeconomic contextual disadvantage. While a nonwhite race is also associated with CMV seropositivity, ADI remains predictive of CMV even in models that adjust for race. ADI percentile is significantly higher among seropositive than seronegative women when each racial category is evaluated independently. This relationship between maternal CMV and ADI suggests that race is merely a marker of socioeconomic disadvantage rather than a CMV risk factor per se. While CMV seroprevalence is spatially variable, this variability largely disappears after adjustment for race and ADI; this suggests that the distribution of CMV is closely related to socioeconomic and demographic factors. Our 23 subjects with recently acquired CMV had an average ADI percentile that was between that of the seropositive and the seropositive women. This may be because some of the recent CMV infections are occurring in the more affluent neighborhoods where proportionally more women are susceptible to CMV.

We do not have a complete understanding as to why CMV disproportionately affects the poor. It is most plausible that certain social factors, including household composition, crowding, contact with children, and socially segregated sexual networks, are associated with CMV risk, and these risks themselves segregate alongside both race and poverty [9, 10]. In our previous study, we found that CMV seropositivity rates were similar among both white and African American children [11]. Beginning in the teenage years, however, African American teenagers had markedly higher seropositivity rates than whites. This suggests that sexual acquisition may be an important driver of the excess CMV exposure among socially disadvantaged populations [16].

Electronic medical records have greatly facilitated retrospective studies using large patient cohorts. While considerable clinical detail can be populated from electronic medical records, demographic data is usually limited to age, gender, and self-reported race and ethnicity. These variables shed little light on important social determinants of health, such as education levels, family structure, income, and the built environment. While individual demographic data can be sparse, there are abundant data available for spatial units such as census block groups. Thus, having geographic identifiers for patients allows us to evaluate relationships between *neighborhood variables* and *individual health outcomes*. Choosing from among the hundreds of demographic variables in the US Census and the American Community Survey presents its own challenges, as many variables are highly correlated, and because it is seldom clear which variables among these are the best with which to populate a statistical model. The ADI provides several advantages as it (1) is a composite of 17 individual variables, (2) does not contain any direct health measures, (3) has a growing literature basis associating it with health outcomes, and (4) is freely available through the University of Wisconsin for others to use [14].

Our study has several important limitations. The coordinate data for our study subjects represents their reported address,
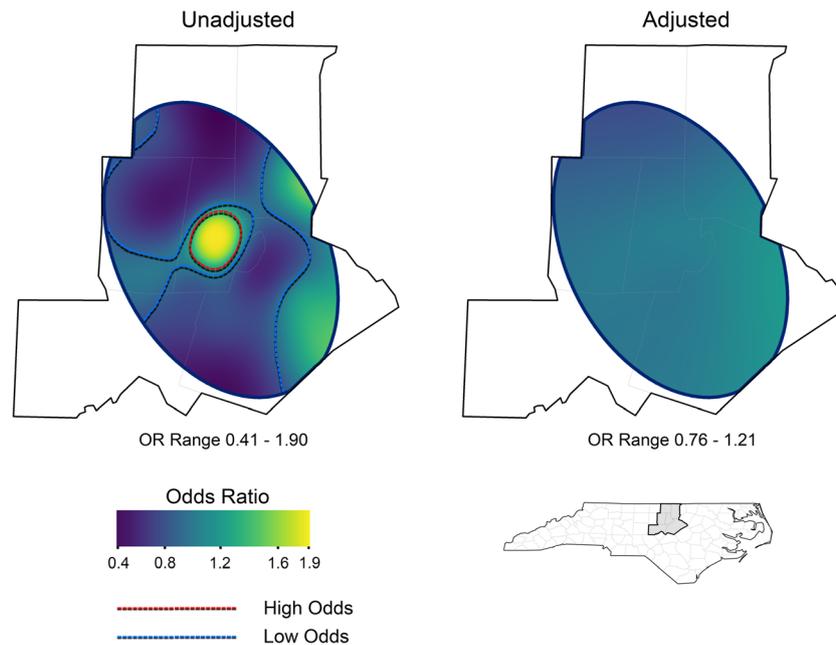
Fig. 1 Generalized additive models showing the geographic heterogeneity of maternal CMV seropositivity. In the unadjusted model, maternal CMV is compared only to smoothed effects of longitude and latitude. The local odds ratio in the unadjusted model varied from 0.41 to 1.90 compared with the average odds. The odds of CMV seropositivity were significantly higher than average in the urban neighborhoods of Durham, while they were significantly low in the more affluent suburbs. The adjusted model included both neighborhood-level ADI and an interaction term for individual age and race. This adjustment substantially blunted the odds ratio range (0.76 to 1.21) and effaced much of the geographic variability of CMV odds. Thus, the combination of neighborhood ADI and individual age and race statistically explains much of the distribution of CMV seropositivity

but these data may be inaccurate. CMV seropositive women did not necessarily acquire CMV while living at the coordinates available to our study. We have created a binary category of non-Hispanic white versus minority women. It is important to recognize that these categories are themselves based on self-reported race and ethnicity, and in themselves each category may represent diverse demographic subgroups. Our most important statistical limitation is the association of neighborhood-level ADI with individual-level outcomes. Our 3504 subjects are nested within 540 block groups, which means that the same ADI value will be repeated in each individual from a given neighborhood. This may violate a statistical assumption of independence. Moreover, the partitioning of ADI values by neighborhood boundaries is subject to the "modifiable areal unit problem," a source of bias caused by the averaging of values within an arbitrary set of boundaries. Despite these concerns, area-level ADI values appear to autocorrelate, with poor neighborhoods generally adjoining other poor neighborhoods and vice versa. This suggests that ADI represents a spatially continuous trend that is not intrinsically tied to neighborhood boundaries.

To conclude, we have shown that the ADI, a neighborhood-level index of socioeconomic contextual disadvantage, is significantly associated with individual CMV infection among pregnant women. Because the ADI datasets are freely available for the entire USA, ADI can potentially serve as a valuable tool for identifying neighborhoods with a high prevalence of CMV and other health states associated with socioeconomic disadvantage. Further research is needed to identify the social or biological determinants of CMV risk among women in poor communities. However, understanding the geographically and demographically disparate impact of CMV may be valuable for educating community members, as well as for modeling the utility of maternal and newborn screening in these communities.

Compliance with Ethical Standards

Conflicts of Interest The authors declare that they have no conflicts of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964

Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the Duke University Institutional Review Board. A waiver of Informed consent was granted for this retrospective study. Measures were taken to secure confidential subject data in compliance with HIPAA and with standards for confidential patient data by Duke University School of Medicine and Duke Health Technology Solutions.

References

1. Cannon MJ, Davis KF. Washing our hands of the congenital cytomegalovirus disease epidemic. BMC Public Health. 2005;5:70.
2. Colugnati FA, Staras SA, Dollard SC, Cannon MJ. Incidence of cytomegalovirus infection among the general population and pregnant women in the United States. BMC Infect Dis. 2007;7:71.
3. Stadler LP, Bernstein DI, Callahan ST, Turley CB, Munoz FM, Ferreira J, et al. Seroprevalence and risk factors for cytomegalovirus infections in adolescent females. J Pediatr Infect Dis Soc. 2013;2(1):7–14.
4. Stadler LP, Bernstein DI, Callahan ST, Ferreira J, Gorgone Simone GA, Edwards KM, et al. Seroprevalence of cytomegalovirus (CMV) and risk factors for infection in adolescent males. Clin Infect Dis. 2010;51(10):e76–81.
5. Sohn YM, Oh MK, Balcarek KB, Cloud GA, Pass RF. Cytomegalovirus infection in sexually active adolescents. J Infect Dis. 1991;163(3):460–3.
6. Gaffey MJ, Tucker RM, Fisch MJ, Normansell DE. The seroprevalence of cytomegalovirus among Virginia State prisoners. Public Health. 1989;103(4):303–6.
7. Bate SL, Dollard SC, Cannon MJ. Cytomegalovirus seroprevalence in the United States: the national health and nutrition examination surveys, 1988-2004. Clin Infect Dis. 2010;50(11):1439–47.
8. Bristow BN, O'Keefe KA, Shafir SC, Sorvillo FJ. Congenital cytomegalovirus mortality in the United States, 1990-2006. PLoS Negl Trop Dis. 2011;5(4):e1140.
9. Dowd JB, Palermo TM, Aiello AE. Family poverty is associated with cytomegalovirus antibody titers in U.S. children. Health Psychol. 2012;31(1):5–10.
10. Griffiths P, Baboonian C, Ashby D. The demographic characteristics of pregnant women infected with cytomegalovirus. Int J Epidemiol. 1985;14(3):447–52.
11. Lantos PM, Permar SR, Hoffman K, Swamy GK. The Excess Burden of Cytomegalovirus in African American Communities: A Geospatial Analysis. Open Forum Infect Dis. 2015;2(4):ofv180.
12. Lantos PM, Hoffman K, Permar SR. Jackson P. Swamy GK. Geographic Disparities in Cytomegalovirus Infection During Pregnancy. J Pediatric Infect Dis Soc: Hughes BL; 2017.
13. The ADI is freely available through the University of Wisconsin by emailing neighborhoodatlas@medicine.wisc.edu.
14. Kind AJ, Jencks S, Brock J, Yu M, Bartels C, Ehlenbach W, et al. Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. Ann Intern Med. 2014;161(11):765–74.
15. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J Royal Stat Soc (B). 2011;73(1):3–36.
16. Staras SA, Flanders WD, Dollard SC, Pass RF, McGowan JE Jr, Cannon MJ. Influence of sexual activity on cytomegalovirus seroprevalence in the United States, 1988-1994. Sex Transm Dis. 2008;35(5):472–9.