MAJOR ARTICLE

# The Excess Burden of Cytomegalovirus in African American Communities: A Geospatial Analysis

**Paul M. Lantos,[1,2] Sallie R. Permar,[2] Kate Hoffman,[4] and Geeta K. Swamy[3]**

Divisions of [1]Pediatric Infectious Diseases; [2]General Internal Medicine; [3]Department of Obstetrics and Gynecology, Duke University School of Medicine, and [4]Nicholas School of the Environment, Duke University, Durham, North Carolina

***Background.*** Cytomegalovirus (CMV) is a common cause of birth defects and hearing loss in infants and opportunistic infections in the immunocompromised. Previous studies have found higher CMV seroprevalence rates among minorities and among persons with lower socioeconomic status. No studies have investigated the geographic distribution of CMV and its relationship to age, race, and poverty in the community.

***Methods.*** We identified patients from 6 North Carolina counties who were tested in the Duke University Health System for CMV immunoglobulin G. We performed spatial statistical analyses to analyze the distributions of seropositive and seronegative individuals.

***Results.*** Of 1884 subjects, 90% were either white or African American. Cytomegalovirus seropositivity was significantly more common among African Americans (73% vs 42%; odds ratio, 3.31; 95% confidence interval, 2.7–4.1), and this disparity persisted across the life span. We identified clusters of high and low CMV odds, both of which were largely explained by race. Clusters of high CMV odds were found in communities with high proportions of African Americans.

***Conclusions.*** Cytomegalovirus seropositivity is geographically clustered, and its distribution is strongly determined by a community's racial composition. African American communities have high prevalence rates of CMV infection, and there may be a disparate burden of CMV-associated morbidity in these communities.

***Keywords.*** African American; cytomegalovirus; disparity; epidemiology; geographic information system.

Cytomegalovirus (CMV) infection is a common community-acquired viral infection that is found throughout the world. It is the leading infectious cause of neurologic deficits and hearing loss in infants, resulting in more long-term pediatric disabilities than Down's syndrome and spina bifida. It is also an important opportunistic pathogen among immunocompromised individuals, and it has been associated with atherosclerosis and age-associated immunosenescence. Previous studies have associated risk of CMV infection with poverty and racial or ethnic minority status.

We have undertaken a study of CMV using geographic information system (GIS) software and spatial statistical analysis. We used this approach to assess whether seropositive and seronegative patients have different geographic distributions, to determine whether other variables with shared geography (such as census-based demographic variables) are associated with the odds of CMV seropositivity. Ultimately, our aim was to identify communities (1) that may have a disproportionate risk of congenital CMV with its associated morbidity or (2) that constitute a community reservoir for CMV transmission. This analysis could inform public health interventions such as maternal and infant CMV screening.

## METHODS

### Design

This was a retrospective cross-sectional case-control study using electronic health records.

Received 4 November 2015; accepted 13 November 2015.

Correspondence: Paul M. Lantos, MD, DUMC 100800, Durham, NC 27710 (paul.lantos@duke.edu).

**Open Forum Infectious Diseases**



DOI: 10.1093/ofid/ofv180

### Patient Records

We performed an electronic query of the Duke University Health System electronic health records to identify patients who had been tested for CMV immunoglobulin (Ig)G. We included patients whose address of residence was in Durham County, North Carolina or 1 of the 5 bordering counties (Wake, Person, Chatham, Orange, and Granville) and who had been tested between January 1, 2010 and December 31, 2013. In most cases, IgG results were reported qualitatively (eg, positive, negative, or equivocal/indeterminate). In a minority of cases (under 5%), cases were reported as titers with reference ranges. For the purposes of our study, we included patients who could be classified as positive or negative, and we excluded those with only an equivocal or indeterminate result. Patients with multiple tests were classified as positive if they had tested positive once during the 4-year study period. We did not evaluate CMV IgM results, because we were interested in background seroprevalence; IgM is primarily used to document acute infection.

### Geographic Data Management

Residential addresses are geocoded within the Duke informatics system, allowing us to obtain the latitude and longitude coordinates of each patient's residence. These records were projected as point data in a GIS using ArcGIS 10.3 (ESRI, Redlands, CA). Each patient record was joined with select block group-level attributes from the US Census and American Community Survey.

### Statistical Analyses

All nonspatial statistics were performed using the statistical programming language R (www.cran-project.org), including multivariate logistic regression, *t* tests, Boschloo's exact test [1,2], and Fisher's exact test. We used ArcGIS for descriptive spatial statistics, including calculation of standard deviational ellipses. We used the spatstat and generalized additive model (GAM) packages in R for further spatial statistical analyses.

We performed the Ripley's cross-L function for marked point data to determine whether different point patterns deviated from complete spatial randomness. We calculated the statistical significance of Ripley's cross-L function using a test of maximum absolute deviation from complete spatial randomness and the Diggle-Gressie-Loosmore-Ford test, each with 9999 Monte Carlo simulations.

Next, we used a GAM to calculate odds ratio (OR) surfaces and clustering of CMV seropositivity. This was accomplished using the GAM package in R [3]. The GAM analysis is similar to logistic regression, but it incorporates a 2-dimensional spatial smoothing factor using local regression ("loess"). To predict ORs, the loess smoother utilizes information from nearby data points, weighting information based on its distance from the prediction point. The region or neighborhood from which data are drawn to predict prevalence is based on the percentage of data points in the neighborhood; this is referred to as the span size. The optimal span size for the loess function is that which minimizes the Aikake's Information Criterion. We performed an unadjusted model, followed by an adjusted model incorporating patient race. A span size of 0.15 was used for the unadjusted model and 0.50 for the adjusted model. Our initial mapping of the 6 study counties revealed outlying areas with sparse numbers of subjects. To avoid making predictions in areas with sparse data, we calculated an ellipse containing 2 standard deviations of patients, then we clipped this ellipse to the boundaries of the 6 counties.

### Ethical Review

This study was approved by the Duke University Institutional Review Board. A waiver of informed consent was approved for this retrospective study.

## RESULTS

### Query Results

We identified 1884 unique patients who had been serologically tested for CMV during the 4-year recruitment period. There were 264 subjects who had been tested at least twice. Eleven of the multiply tested patients had discordant results with at least 1 positive CMV IgG test, and these individuals were categorized as seropositive. Ninety percent of our cohort was either African American (612, 32%) or white (1087, 58%). The remaining 10% was a mix of other self-identified races (53 Asian, 38 multiracial, 2 American Indian, 1 Hawaii-Pacific, 74 "other," and 17 declined or unavailable). Fifty-eight subjects self-identified as Hispanic or Latino, 44 of whom listed their race as "other." We conducted all further analyses only on the African American and white subjects. All neighborhood-level means demonstrated significantly worse indices of poverty and crowding among African Americans compared with whites (Table 1).

**Table 1. Neighborhood (Census Block Group) Means for 1533 Subjects Whose Residential Coordinates Were Known[a]**

| Units[b] | African American | White | *P* Value |
|---|---|---|---|
| Average Household Size | 2.54 | 2.47 | .0003 |
| Poverty | 18.2 | 8.9 | <.0001 |
| Median Household Income | 51 859 | 75 370 | <.0001 |
| Unemployment | 6.86 | 4.27 | <.0001 |
| High School Graduates | 84.9 | 92.2 | <.0001 |
| Population Density | 2512 | 1763 | <.0001 |

[a] All comparisons were performed using Welch's 2-sample *t* test. Data were obtained from the 2010 United States Census and the 2013 American Community Survey.

[b] Average Household Size = persons; Poverty = percentage of population below the poverty level; Median Household Income = dollars; Unemployment = percentage of total adult population unemployed; High School Graduates = percentage of the population that has completed high school or equivalent; Population Density = persons per square mile.

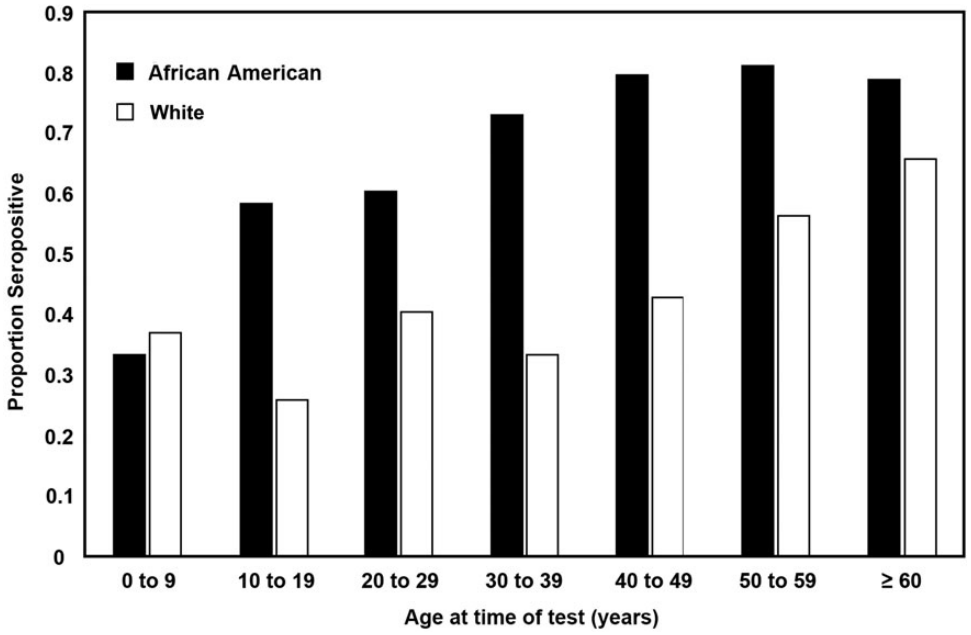


**Figure 1.** Age distribution of cytomegalovirus (CMV) seropositive and seronegative individuals. Both African American and white children had similar rates of CMV seropositivity. Beyond age 10, however, African Americans had a seroprevalence approximately twice that of whites. Only among the older age cohorts did the CMV seroprevalence among whites approach that of African Americans.

Overall, 1060 subjects tested positive and 796 tested negative for CMV IgG (57.1%, 95% CI, 54.8%–59.3%). Equivocal or indeterminate results had been reported for 28 subjects, or 1.5% of the total cohort. These subjects were excluded from further analysis. Cytomegalovirus seropositivity was significantly more prevalent among African Americans than whites (73% vs 42%; OR = 3.31; 95% confidence interval [CI], 2.7–4.1). Children who were tested before 10 years of age had similar rates of seropositivity in both groups (Figure 1). In contrast, among 10- to 19-year-olds, African Americans had a CMV seroprevalence more than twice that of their white counterparts (58% vs 26%; OR = 3.99; 95% CI, 1.9–8.7), and this discrepancy remained true through age 50. Only among individuals 60 years of age and older were the seroprevalence rates similar among races.

### Spatial Analyses

Of the 1884 subjects identified in our records search, geographic coordinates were available for 1856.

We calculated 1 standard deviation ellipses for CMV-seropositive and CMV-seronegative patients in each racial group. The seropositive and seronegative ellipses were similar to one another when each racial group was analyzed individually (Figure 2). This suggested that the distribution of CMV seropositivity primarily reflected the overall distribution of that racial group rather than an inherently different distribution of seropositive and seronegative individuals. This was confirmed by Ripley's cross-L function (Figure 3), in which there was significant ($P < .001$) spatial dependence between seropositive and seronegative subjects at all distances. In other words, the seropositive and seronegative individuals were spatially clustered beyond what would be expected with complete spatial randomness.

The GAM was performed using an ellipse that contained 1533 total subjects. The geostatistical surface (Figure 4A) produced by our unadjusted model had an OR range from 0.4 to 3.2. Two significant clusters of high OR were identified, corresponding to the cities of Durham and Raleigh. These high OR clusters contained 57 (5.8%) whites and 182 (33%) African Americans in the study cohort. African Americans had significantly higher odds of being CMV seropositive if they lived within these clusters (81% vs 59%; OR = 1.92; 95% CI, 1.22–3.06; $P < .003$). Whites exhibited a nonsignificant trend towards higher CMV seropositivity rates within these clusters as well (54% vs 44%; OR = 1.51; 95% CI, .85–2.70; $P = .13$). There were also 4 significant clusters of lower OR. Adjusting our model for subject race eliminated 3 of these low OR clusters, diminished the size of the remaining clusters, and blunted the OR range to 0.5–2.1 (Figure 4B). The high OR clusters largely encircled neighborhoods with a high proportion of African Americans in the general population; the opposite was true of the low OR clusters (Figure 4C).

## DISCUSSION

We have found that CMV-seropositive individuals are significantly clustered within urban areas with high minority populations and poor socioeconomic indicators. This raises the

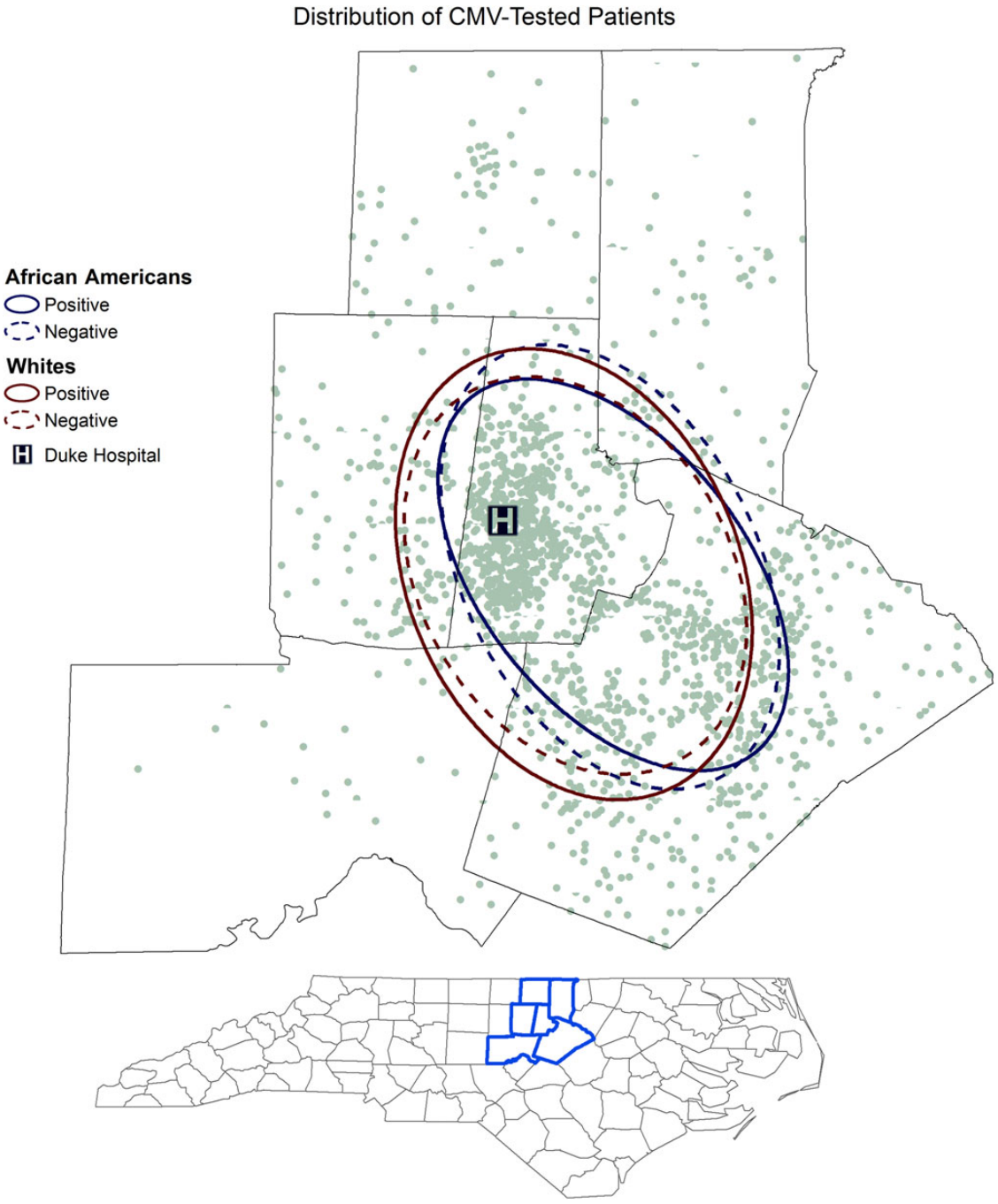


**Figure 2.** Spatial distribution of patients tested for cytomegalovirus (CMV). Ellipses contain 1 standard deviation of both seropositive and seronegative African Americans and whites. Patients are represented by dots placed randomly within their census block group (the boundaries of which are not shown). Although there is considerable overlap of all ellipses, the spatial distribution is more closely aligned with race than with test result.

question of whether CMV-attributable morbidity, particularly the neurologic sequelae of congenital infection, are also concentrated in these areas. This may add to a public health burden experienced by entire communities that are already underserved. Moreover, the high seroprevalence in these communities may constitute a reservoir for this pathogen, affecting the surrounding geographic areas as well.

Among patients tested at our institution, African Americans were nearly twice as likely to be CMV seropositive as whites. Adjusting for race largely abrogated clustering, suggesting that the distribution of CMV-seropositive individuals is related to the spatial segregation of African Americans and whites. This was confirmed by superimposing the boundaries of the clusters on a census map, revealing that high odds clusters corresponded

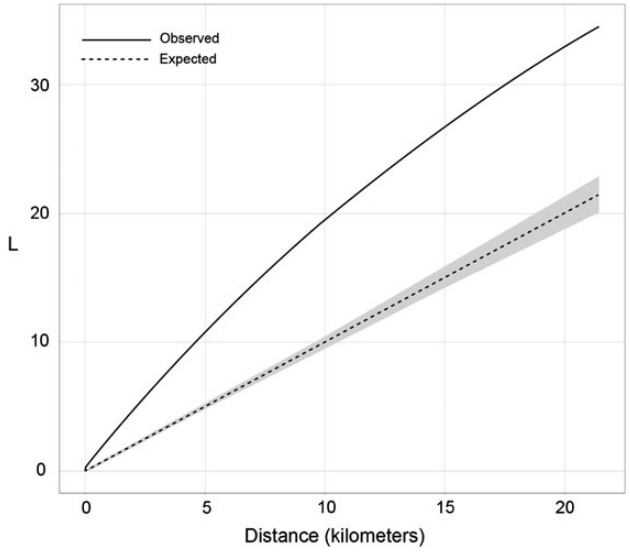


**Figure 3.** Ripley's cross-L function: seropositive versus seronegative point pattern analysis. This statistical test determines whether 2 point patterns are independent or clustered compared with complete spatial randomness. The "Expected" line is that which would be observed under conditions of complete spatial randomness, and the envelope is its 95% confidence interval. The horizontal axis represents distance between any 2 points. The vertical axis represents the calculated L function; values above the spatial randomness envelope represent significant clustering (ie, spatial association). In this case, the relationship between seropositive and seronegative spatial patterns significantly departed from complete spatial randomness (maximum absolute deviation from complete spatial randomness test and Diggle-Gressie-Loosmore-Ford test, both $P < .001$). This suggests that at a global level seropositive and seronegative individuals do not have significantly different spatial patterns from one another.

to neighborhoods with a high proportion of African Americans. The opposite was true for low odds clusters. We then must conclude that the burden of CMV is experienced disproportionately not only among African American individuals, but in neighborhoods and communities with large African American populations. This excess was demonstrable among adolescent and young adult subjects, overlapping the age of fertility. This may be responsible for to the high risk of congenital CMV transmission that has been documented among young African Americans [4, 5].

Our findings corroborate other studies in which racial and ethnic minorities have substantially higher CMV seropositivity rates than whites. This has been true among diverse patient populations and age groups, including pregnant women [6], adolescents [7–9], male prisoners [10], and the general US population [6, 11]. African American and Native American infants suffer a disproportionate burden of congenital CMV mortality [12]. Cytomegalovirus seropositivity has also been associated with poverty [13, 14].

What is not discernable from our study is why CMV is so much more prevalent among African Americans than whites. Cytomegalovirus is transmitted among humans by numerous mechanisms, including congenital transmission, oral secretions, breastfeeding, sexual contact, fomites, and transfusion or transplantation. If urban poverty creates the conditions for high rates of CMV, it is worth entertaining 3 explanations. First, do urban minority communities have more children in daycare or group infant care settings, or in daycares that are more crowded—and as a consequence are adults exposed to CMV primarily through their children? Daycare workers and children in daycare (versus home care) are at increased risk of acquiring CMV [15, 16]. Some evidence suggests that parents of children in daycare are also at increased risk [17, 18]. We find this insufficient to account for the large racial disparity, because in our dataset the CMV rates among children under 10 years of age was similar in both African Americans and whites. Another potential explanation is crowding in the home environment, including multiple family members sharing bedrooms or conditions of impoverished hygiene. Our data do not shed light on this possibility: we only had access to neighborhood level statistics such as average household size; it would require interviews with individual patients to identify a relationship between household conditions and CMV serostatus. Finally, sexual transmission of CMV may be the critical explanation. Cervical secretion of CMV is associated with sexual activity, CMV acquisition is more common among patients with other sexually transmitted infections, and condom use is protective against CMV acquisition among men [9, 19–25]. Recent onset of maternal sexual activity and maternal sexually transmitted infections have been associated with congenital CMV infection [4, 5]. Sexually transmitted CMV may be unto itself a socially complex phenomenon, incorporating variables such as age of sexual debut, number of sexual partners, use of barrier protective methods, and socially segregated sexual networks.

There are limitations and potential biases in this retrospective study. An important limitation was our inability to identify clinical subgroups within our cohort. Cytomegalovirus testing is not routinely performed on the general, healthy population. It is most often performed in settings for patients with organ transplantation, malignancies, human immunodeficiency virus/acquired immune deficiency syndrome, and suspected congenital infection; it is also occasionally performed to evaluate syndromes such as mononucleosis-like illnesses, hepatitis, and colitis. It is likely that many subjects in our study cohort could be grouped accordingly. Similarly, we do not know whether African American and white patients were tested for the same indications and in the same proportions. At the same time, our data show such a marked disparity between these 2 groups that it is unlikely to be solely an artifact of testing biases.

An important bias in spatial epidemiology research occurs when the disease population and control population are identified using different sampling methods. This could, for instance, create the appearance of clustering that is better explained by heterogeneities in population density. We were able to avoid

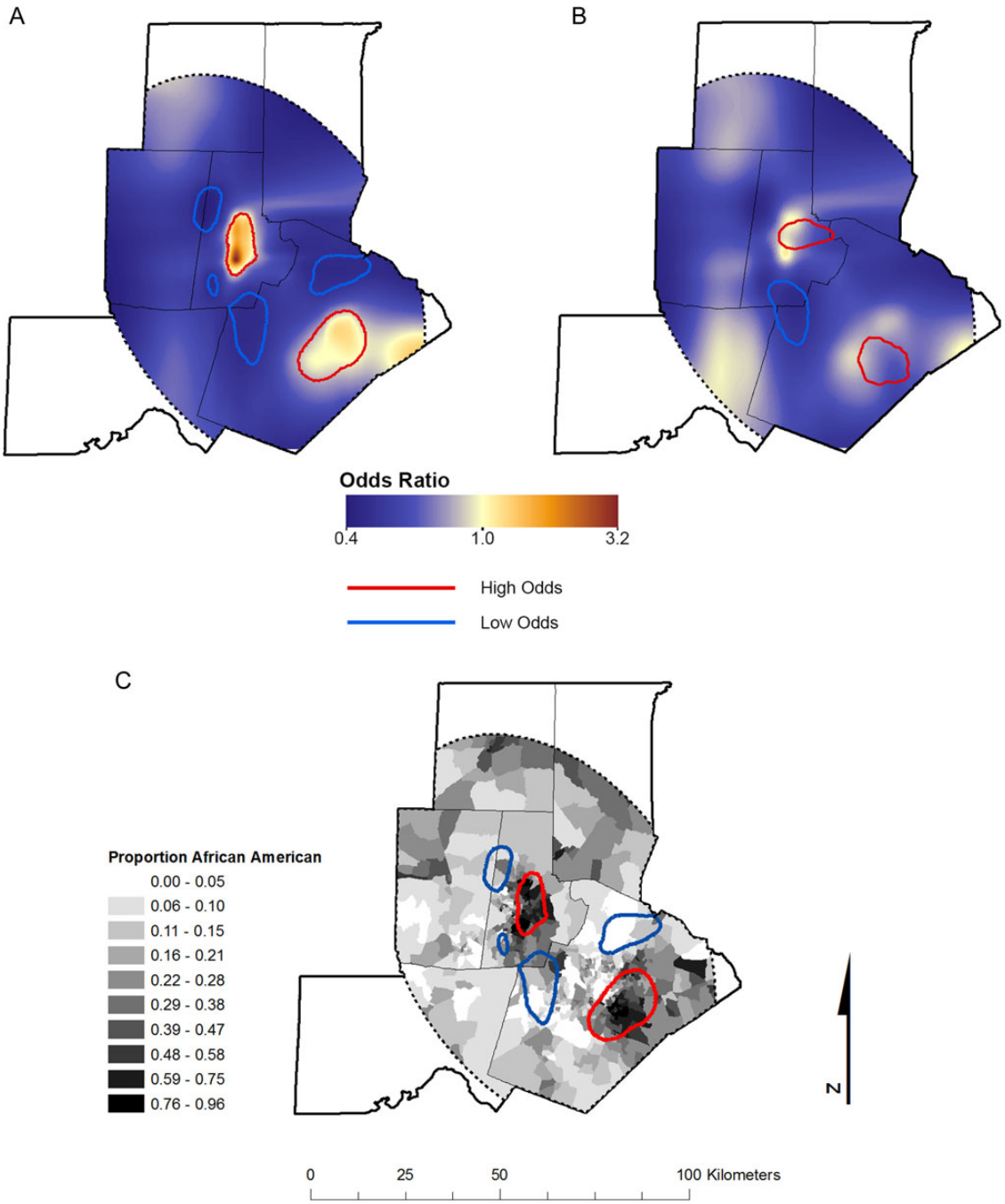


**Figure 4.** Generalized additive model. This analysis is a logistic regression with a 2-dimensional spatial smoothing function, resulting in a continuous odds ratio (OR) surface. The response variable in this model was the binary result of cytomegalovirus (CMV) testing. Significant deviations from an OR of 1 are encircled with a contour, representing a 2-tailed *P* value of .05. (A) An unadjusted model identified clusters with significantly high and significantly low odds of CMV seropositivity. The OR range was 0.4–3.2. (B) Adjusting for race diminished the OR range, eliminated all but one of the low risk clusters, and diminished the size of the remaining ones. This indicates that race is a key predictor of local odds of CMV seropositivity. (C) Superimposing the contours on a census map shows that areas of significantly high CMV risk are those with a high proportion of African Americans. By contrast, areas of significantly low CMV risk have a low proportion of African Americans. The 2 high odds clusters surround the cities of Durham and Raleigh, North Carolina.

this bias by using a case-control study design. We sampled all of our study subjects, irrespective of test result, using a single query. Thus, we were able to statistically test the null hypothesis that positive and negative control subjects had a random spatial relationship with one another. Therefore, the observation of clustering within 2 urban centers demonstrates excess cases with respect to the control distribution, rather than a concentration of cases in an area with high population density.

## CONCLUSIONS

More research is clearly needed to quantify the burden of CMV disease among minorities, rather than mere seropositivity. We do not yet understand whether sexual transmission is the primary driver of CMV acquisition in minority communities, and, if so, whether promoting safe sexual practices could diminish CMV seroprevalence and thereby reduce CMV morbidity in infant and immunosuppressed populations. We also need a more complete picture of which demographic and socioeconomic factors, whether at the individual or the community level, can predict risk of CMV. This greater understanding of population factors on CMV acquisition rates will benefit ongoing models of CMV vaccination implementation and prenatal and infant screening.


## Acknowledgments

We thank Dr. Jason Stout for statistical advice and support.

***Disclaimer.*** The contents of this article are solely the responsibility of the authors and do not necessarily represent the official view of the National Institutes of Health.

***Financial support.*** P. M. L. was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH) under award number KL2 TR001115. S. R. P. was supported by the NIH Director's New Innovator Award DP2 grant, number HD075699.

***Potential conflicts of interest.*** S. R. P. has a nonfinancial consulting agreement with Pfizer Vaccines.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.


## References


1. Calhoun P. Unconditional Exact Test. Available at: https://cran.r-project.org/web/packages/Exact/Exact.pdf. Accessed 1 September 2015.
2. Boschloo R. Raised conditional level of significance for the 2 × 2-table when testing the equality of two probabiities. Statistica Neerlandica **1970**; 24:1–35.
3. Vieira V, Bartell S, Bliss R. Mapping Smoothed Effect Estimates from Individual-Level Data. Available at: https://cran.r-project.org/web/packages/MapGAM/MapGAM.pdf. Accessed 1 September 2015.
4. Fowler KB, Pass RF. Sexually transmitted diseases in mothers of neonates with congenital cytomegalovirus infection. J Infect Dis **1991**; 164:259–64.
5. Fowler KB, Pass RF. Risk factors for congenital cytomegalovirus infection in the offspring of young women: exposure to young children and recent onset of sexual activity. Pediatrics **2006**; 118:e286–92.
6. 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.
7. Stadler LP, Bernstein DI, Callahan ST, et al. Seroprevalence and risk factors for cytomegalovirus infections in adolescent females. J Pediatric Infect Dis Soc **2013**; 2:7–14.
8. Stadler LP, Bernstein DI, Callahan ST, et al. Seroprevalence of cytomegalovirus (CMV) and risk factors for infection in adolescent males. Clin Infect Dis **2010**; 51:e76–81.
9. Sohn YM, Oh MK, Balcarek KB, et al. Cytomegalovirus infection in sexually active adolescents. J Infect Dis **1991**; 163:460–3.
10. Gaffey MJ, Tucker RM, Fisch MJ, Normansell DE. The seroprevalence of cytomegalovirus among Virginia State prisoners. Public Health **1989**; 103:303–6.
11. 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**; 50:1439–47.
12. Bristow BN, O'Keefe KA, Shafir SC, Sorvillo FJ. Congenital cytomegalovirus mortality in the United States, 1990–2006. PLoS Negl Trop Dis **2011**; 5:e1140.
13. Dowd JB, Palermo TM, Aiello AE. Family poverty is associated with cytomegalovirus antibody titers in U.S. children. Health Psychol **2012**; 31:5–10.
14. Griffiths P, Baboonian C, Ashby D. The demographic characteristics of pregnant women infected with cytomegalovirus. Int J Epidemiol **1985**; 14:447–52.
15. Joseph SA, Beliveau C, Muecke CJ, et al. Cytomegalovirus as an occupational risk in daycare educators. Paediatr Child Health **2006**; 11:401–7.
16. Hutto C, Ricks R, Garvie M, Pass RF. Epidemiology of cytomegalovirus infections in young children: day care vs. home care. Pediatr Infect Dis **1985**; 4:149–52.
17. Adler SP. Cytomegalovirus and child day care: risk factors for maternal infection. Pediatr Infect Dis J **1991**; 10:590–4.
18. Pass RF, Hutto C, Ricks R, Cloud GA. Increased rate of cytomegalovirus infection among parents of children attending day-care centers. N Engl J Med **1986**; 314:1414–8.
19. Carre N, Robain M, Dussaix E, et al. Sexual factors associated with cytomegalovirus seropositivity in human immunodeficiency virus-infected men. The Seroco Study Group. Sex Transm Dis **1997**; 24:582–6.
20. Collier AC, Handsfield HH, Ashley R, et al. Cervical but not urinary excretion of cytomegalovirus is related to sexual activity and contraceptive practices in sexually active women. J Infect Dis **1995**; 171:33–8.
21. Shen CY, Chang SF, Lin HJ, et al. Cervical cytomegalovirus infection in prostitutes and in women attending a sexually transmitted disease clinic. J Med Virol **1994**; 43:362–6.
22. Collier AC, Handsfield HH, Roberts PL, et al. Cytomegalovirus infection in women attending a sexually transmitted disease clinic. J Infect Dis **1990**; 162:46–51.
23. Handsfield HH, Chandler SH, Caine VA, et al. Cytomegalovirus infection in sex partners: evidence for sexual transmission. J Infect Dis **1985**; 151:344–8.
24. Chandler SH, Holmes KK, Wentworth BB, et al. The epidemiology of cytomegaloviral infection in women attending a sexually transmitted disease clinic. J Infect Dis **1985**; 152:597–605.
25. Davis LE, Stewart JA, Garvin S. Cytomegalovirus infection: a seroepidemiologic comparison of nuns and women from a venereal disease clinic. Am J Epidemiol **1975**; 102:327–30.