Geospatial Mapping & Analysis of Health Care Conditions in Children

Brian R. Jacobs, MD  
VP and Chief Medical Information Officer  
Children’s National Medical Center  
Washington, DC

Background:
Childhood obesity prevalence data is based largely on survey data, with indications that > 17% of children are obese in the United States. The electronic health record (EHR) provides a rich source of accurate and reliable body mass index (BMI) data. The purpose of this study was to investigate the prevalence of childhood obesity in the Washington, DC metropolitan area using EHR-derived data and to examine the differences between 3 distinct populations – inner city DC, suburban Maryland and hospitalized patients. An additional goal was to examine the relationship between race and BMI in this region.

Leading healthcare organizations have replaced paper-based medical records with electronic health records. The EHR provides a rich source of electronically accessible discrete patient data around demographics, race, age, patient address, vital signs, laboratory results, radiology results and healthcare provider clinical documentation. As this detailed data are electronically available, there is an opportunity to aggregate, analyze and compare its geographic characteristics to related regional co-variables utilizing geospatial information systems (GIS) technology. The use of GIS in the evaluation of EHR-derived data is of major significance to investigators as it permits geographic and co-factor specific targeting of epidemiologic methods, and, preventive and therapeutic treatment trials for patients with various conditions. Children’s National Medical Center has initiated a partnership with the following stakeholders: Children’s National faculty members and the George Washington University Department of Geography’s GIS team have joined together to establish the foundation to improve childhood health in the DC metropolitan region (District of Columbia, Maryland and Virginia). This project took advantage of essential health data derived from 3 regional EHRs: The Children’s National inpatient EHR, the Goldberg Center for Primary Pediatric Care EHR (7 independent inner city clinics and mobile vans) and the Children’s Pediatricians and Associates (CP&A) EHR representing 10 independent suburban pediatric practices. These 3 EHRs contain over 350,000 health encounters and close to 50,000 unique patient records each year representing care delivery for children in the DC Metropolitan region.

In any given population, there are socioeconomic and environmental characteristics which influence healthcare condition prevalence. By deriving patient characteristics from the EHRs within a region and combining this information with other known geospatial data sets (such as the US Census), investigators can easily highlight high density problem foci. Furthermore, co-variables associated or postulated to be associated with these problem foci
may be clearly identified and aligned with that population. Study design, implementation, and recruitment can therefore be more targeted.

To date, 3 pilot healthcare geospatial projects have been carried out in the DC metropolitan region. In each project, a specific health condition was targeted, patient-specific data aggregated and analyzed, GIS technologies utilized, and study results disseminated. The first project focused on childhood immunization, the second on thermal burns and in the current project we focus on childhood obesity.

**Geospatial Relationships of Childhood Immunizations:** In this study our group examined the relationship between spatial accessibility to pediatric immunization providers and vaccination compliance in a low income, urban population of children. We used electronic data from the 2007 Washington, DC, Immunization Information System (IIS) to collect data on the immunization statuses and residential addresses of Medicaid insured children who were aged 19 to 35 months. As part of the methodology we calculated the spatial accessibility to pediatric vaccination providers for each child by assessing the provider-to-population ratio at each residential address. The relationship between spatial accessibility to providers and vaccination compliance was examined. Overall for our cohort of 4,195 children, 80.5% of the children were up-to-date with vaccinations. Vaccination coverage ranged from 61.6% to 100% (median: 79.2%) among different neighborhoods. Having the highest level of access to pediatric vaccination providers was associated with 36% higher odds of being up-to-date as compared with having the lowest level of access (Figure 1). We concluded that children with greater spatial accessibility to pediatric vaccination providers were more likely to be up-to-date with vaccinations.

**Geospatial relationships of childhood thermal burns:** Burn injuries are a leading cause of pediatric morbidity and mortality. To reduce the health burden of these injuries, injury prevention strategies that target high-risk populations are needed. In this study, we set out to identify areas in the District of Columbia with an increased number of pediatric burn injuries and to identify demographic and geographic subgroups at risk for these injuries. We utilized

![Figure 1. Relationship of pediatric primary vaccine completion to vaccine provider availability in the DC metropolitan region.](image-url)
electronic demographic and diagnosis data on children <15 years old treated at Children’s National for thermal burns between 2007 and 2009. GIS analysis was used to identify neighborhoods with the highest concentrations of burn injuries and identify the characteristics of injured children in these areas. We found that burn injuries accounted for 107 hospital admissions and 1,713 emergency department and outpatient clinic visits. Children from 47 of the 69 District of Columbia neighborhoods were treated for burn injuries at Children’s. The number of burns within each neighborhood ranged from one to 67. Almost half of these injured children lived in one of six neighborhoods (Figure 2). Within these six neighborhoods, variations in race, insurance status, mechanism of injury, and season of injury were observed. This study shows the value of geographic mapping for defining areas with the highest rate of burn injury at the neighborhood level. Using this approach, burn prevention efforts can be more tailored to populations at risk, potentially reducing overall cost while maintaining effectiveness.

Using the results from this study, we have applied several different strategies for burn prevention. In recognition of the frequency of burn injury in Hispanic toddlers, we partnered with political and advocacy groups including the District of Columbia Mayor’s Office on Latino Affairs to better reach the Spanish-speaking community. In this collaborative effort, we developed single page handouts in Spanish that described simple prevention tips for the most common burn mechanisms in Latino children.

Figure 2. Distribution of burn cases by neighborhood in Washington DC. Darker shaded areas represent higher numbers of burn victims.

The primary goal of the current research initiative was to advance a novel methodology to greatly enhance the practice of medicine and conduct of outcomes research using EHR-derived data. We set out to demonstrate the value of combining EHR-derived data with geospatial mapping tools in understanding childhood obesity within the region. The secondary goal of the project was to advance analytical methods through the incorporation of mixed methods research design, including geospatial analysis of EHR-derived patient data, which is irregularly spaced, multivariate, and longitudinal in nature.

Methods:
Over a 1-year period, data derived from the EHR of Children’s National inpatients, DC ambulatory patients and Maryland suburban patients were extracted and analyzed for unique patients between ages 2 and 20. Gender, height,
weight and age were used to calculate BMI percentiles using CDC formulas. Children were categorized as underweight (<5th percentile), healthy (5th to 85th percentile), overweight (85th to 95th percentile), obese (95th to 99th percentile) or severely obese (>99th percentile).

**Data Collection and Privacy:** Retrospective data was queried from the EHRs of Children’s National, The Goldberg Center for Primary Care Pediatrics (an inner-city ambulatory practice), and Children’s Pediatricians and Associates (CP&A – a consortium of suburban ambulatory practices) to include all children from birth through 21 years of age. These practices in total consist of over 450 healthcare providers and over 350,000 patient encounters each year. Protected Health Information (PHI) resided on one secure computing device which was password protected and accessible to only one of the co-investigators. Subsequent data sets post analysis were stripped of all PHI.

**Protection of Human Subjects:** The research methodology included only existing retrospective data obtained from EHRs in both inpatient and ambulatory environments. No data was collected specifically for research purposes. Because these data were retrospective, no subjects were recruited nor consented for the study. All data were de-identified for the 18 PHI variables. Subjects involved in this study did not experience any physical, psychological, social, or legal risks. Patient Health Information (PHI) including name, birth date, medical record number and address were obtained on initial query only of the EHRs. This data was only available on a single password protected computing device, and only the PI and Co-Investigators had access. All PHI was subsequently deleted or converted as described above to non-identifiable information.

**GIS Methodology:** Geographical Information Systems (GIS) links location based information (spatial) to database information (tabular) enabling the user to visualize patterns, relationships and trends. This means of analysis grants a new perspective to information, which is practically absent from exclusively tabular data. Using GIS, we can manage, analyze, query and interact with geographically referenced information using Spatial Analysis techniques. Spatial Analysis is the study of the distribution and clustering of events and/or objects in space, in conjunction with their attribute characteristics. Spatial Analysis has become increasingly important for health based and epidemiological investigations in identification and characterization of disease clusters; investigations into the cause and/or effects of environmental exposure; association of health outcomes with socio-economic disparities; and also, the evaluation (and monitoring) of community based interventions. Stakeholders from the Geography Department at GWU processed all EHR-derived geospatial datasets for this study. They used corresponding cultural and environmental data sets to create geospatial composite maps illustrating absolute healthcare condition data, geography and associated relevant factors.

**EHR data for Geospatial Analysis & Visualization:** Patient address data derived from EHRs, was de-identified, and geocoded (latitude and longitude determined) to produce GIS point data representing the spatial location of individuals in the original dataset.
**The Analysis Platform:** ArcGIS is a suite of desktop software’s for visualizing (mapping), managing, editing and analyzing data, data which are geospatially referenced. ArcMap is the primary software in the ArcGIS suite, the component used to generate individual geospatial layers and compile maps from that information. A map in GIS terms consists of multiple individual layers of geospatial information. These “layers” take the form of vector or raster data types. Vector data are the simple representation of spatial information using coordinate geometry; points (x,y coordinates), lines, and polygons (shapes, areas, boundaries). Raster data on the other hand is slightly more complex, and represents values for a particular phenomenon as a continuous grid based surface. Based on the two data types specific to the GIS environment, we utilized two methods for the representation of EHR data: vector-based or raster-based layers.

**Results:**

In **Figure 3**, approximately 50,000 unique patient records from 2010 were geographically aggregated to DC political (Wards) boundaries and analyzed, suggesting that approximately 33% of children aged 2 to 20 in the DC metropolitan region are overweight, obese or severely obese. Approximately one-third of patients had a BMI percentile corresponding to the overweight, obese or severely obese categories.

![Figure 3: Distribution of Body Mass Index (BMI) amongst approximately 50,000 children ages 2 to 20 years in the DC Metropolitan region.](image)

There were significant differences between the three study populations (inpatient, suburban and inner city) with a greater prevalence of underweight children in the inpatient group and a greater prevalence of overweight, obese and severely obese children in the inner city population (**Table**).

<table>
<thead>
<tr>
<th>BMI Category</th>
<th>Inpatient</th>
<th>Suburb Outpatient</th>
<th>Inner City Outpatient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underweight</td>
<td>497 (9.2%)*</td>
<td>1,138 (4.5%)</td>
<td>546 (3%)</td>
</tr>
<tr>
<td>Healthy</td>
<td>3,000 (55.8%)</td>
<td>16,490 (65.3%)*</td>
<td>10,867 (59.7%)</td>
</tr>
<tr>
<td>Overweight</td>
<td>788 (14.7%)</td>
<td>3,808 (15.1%)</td>
<td>3,050 (16.8%)*</td>
</tr>
<tr>
<td>Obese</td>
<td>684 (12.7%)</td>
<td>2,830 (11.2%)</td>
<td>2,550 (14%)*</td>
</tr>
<tr>
<td>Severely Obese</td>
<td>407 (7.6%)*</td>
<td>992 (3.9%)</td>
<td>1,193 (6.6%)*</td>
</tr>
</tbody>
</table>

* Indicates percent of patients in this BMI category is greater for this care setting than the other 2 care settings combined p<0.05.

In addition, Hispanic patients had BMI percentiles that were on average 15% greater than Caucasian patients.
Co-variables of interest (popular fast food restaurants) were superimposed on a geospatial map creating a visual relationship between regional obesity and fast food restaurants (Figure 4). In this depiction, large numbers of obese and severely obese children are depicted by red/orange areas, while low numbers are depicted by blue areas. Locations of the fast food restaurants are depicted by yellow squares.

**Figure 4.** Distribution of overweight, obese & severely obese children in the DC region with superimposition of fast food restaurants.

**Conclusions:** Multiple genetic, cultural and environmental factors may predispose individuals to the problem of childhood obesity. The regional EHRs in the DC metropolitan region provide a unique opportunity to gather accurate and reliable data related to the condition of obesity and compare this data to cultural and environmental factors to explore geospatial relationships. The EHR can be used to query, aggregate and analyze large samples of accurate, reliable health data for a population. Using this methodology, we were able to establish patterns of childhood obesity in a major metropolitan region. This approach to population health is likely to uncover opportunities for healthcare improvement, disease prevention targets, and treatment strategies. Geospatial mapping tools allow rapid visualization of condition-specific patterns throughout a geographic region with comparison to relevant co-variable data.
References Cited:


