Pediatric Obesity Subgroups from Electronic Health Record Data

Elizabeth A. Campbell, MSPH, Mitchell G. Maltenfort, PhD, Justine Shults PhD, Christopher B. Forrest, MD, PhD, Aaron J. Masino, ME, PhD
1Drexel University, Philadelphia, PA; 2Children’s Hospital of Philadelphia, Philadelphia, PA; 3University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

Introduction
Childhood obesity remains a major challenge in the United States, where approximately 35 percent of children and adolescents had an age- and sex-specific body mass index (BMI) greater than or equal to the 85th percentile per Centers for Disease Control and Prevention (CDC) growth charts. Despite its prevalence, it remains uncertain if childhood obesity represents a single condition or is composed of different phenotypes with possibly different underlying causes. We present a study of electronic health record (EHR) data for which we sought to identify clinically similar subgroups among a population of newly obese pediatric patients. Specifically, we examine whether certain temporal condition patterns associated with childhood obesity incidence tend to cluster together to characterize subgroups of clinically similar patients.

Methods
In a previous study, a novel application of the sequence mining algorithm, SPADE, was developed and implemented on a large retrospective cohort (n=49,594 patients) to identify common condition trajectories surrounding pediatric obesity incidence. Temporally ordered sequences of conditions were identified from patients’ EHRs for healthcare visits in which an obese BMI was first recorded (the index visit), as well as immediately before (pre-index visit) and after (post-index visit). These sequences were then compared to a matched control population of patients with a healthy BMI. Controls were matched by age, sex, and prior healthcare utilization. SPADE identified 80 temporal condition patterns present in at least 1% of cases that were significantly more common among cases (p<0.05, McNemar’s test). In this study, we used Latent Class Analysis (LCA) to identify potential clusters formed by the temporal condition patterns that were significantly more common among obese pediatric patients. The poLCA package was used in R Version 3.6.1 for this analysis. Each patient was assigned to the group for which he/she had the highest probability of membership. High-prevalence diagnoses, defined as those with ≥20% prevalence among patients in a given LCA-identified class, were used to clinically characterize the subgroups. The Institutional Review Board at the Children’s Hospital of Philadelphia (CHOP) approved this research study and waived the requirement for consent.

Results

Table 1. Latent Class Model Development Comparison

<table>
<thead>
<tr>
<th>Number of Classes</th>
<th>AIC</th>
<th>% reduction in AIC</th>
<th>BIC</th>
<th>% reduction in BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>402,634</td>
<td>-</td>
<td>404,015</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>395,335</td>
<td>1.81%</td>
<td>397,179</td>
<td>1.69%</td>
</tr>
<tr>
<td>5</td>
<td>392,236</td>
<td>0.78%</td>
<td>394,543</td>
<td>0.66%</td>
</tr>
<tr>
<td>6</td>
<td>390,577</td>
<td>0.42%</td>
<td>393,347</td>
<td>0.30%</td>
</tr>
<tr>
<td>7</td>
<td>389,175</td>
<td>0.36%</td>
<td>392,408</td>
<td>0.24%</td>
</tr>
<tr>
<td>8</td>
<td>387,606</td>
<td>0.40%</td>
<td>391,302</td>
<td>0.28%</td>
</tr>
</tbody>
</table>

To obtain a clinically meaningful number of patient subgroups, we restricted evaluation to LCA models with 3-8 classes. Table 1 presents the Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for each model, as well as the percent reduction in each indicator’s value between model iterations. AIC and BIC values both declined as the number of classes in the latent class models increased. The model with 8 latent classes had both the lowest AIC and BIC values, and was the final model selected to study clinical subgroups.

We assessed the eight subgroups identified by LCA for clinical similarity (Table 2). Patients in Class 1 had a high prevalence of respiratory and sleep disorders, including sleep apnea and chronic pharyngitis and tonsillitis. Patients in Class 2 had high rates of inflammatory skin conditions, patients in Class 3 had a high prevalence of seizure disorders, and patients in Class 4 had a high prevalence of Asthma. Patients in Class 5 lacked a clear characteristic morbidity pattern – no condition had a prevalence rate above 5%. Finally, patients in Classes 6, 7, and 8 had a high prevalence of gastrointestinal issues, neurodevelopmental disorders, and physical symptoms (e.g. fever and headaches), respectively. Table 3 presents the mean probability that patients categorized in their respective clusters belong in that group. Mean probability of class membership ranged from 70.21% (patient in Class 8) to 89.7% (patients in Class 1).
Table 2. Prevalence Rates of Common Temporal Diagnoses by Patient Subgroup (n= 49,594 patients). The numbers before each diagnosis in a sequence represents the diagnosis timing class: ‘1’ denotes that the observation was recorded during a patient’s pre-index visit, ‘2’ represents the index visit, and ‘3’ signifies the post-index visit.

<table>
<thead>
<tr>
<th>Class (N)</th>
<th>Diagnoses</th>
</tr>
</thead>
</table>
| Class 1 (n = 2,336): Upper Respiratory and Sleep Disorders | 1-Chronic pharyngitis and tonsillitis (33.05%)  
1-Sleep Apnea (23.54%)  
1-Sleep Problems (21.79%)  
2-Chronic pharyngitis /tonsillitis (45.51%)  
2-Sleep Apnea (26.5%)  
2-Sleep Problems (22.35%)  
3-Chronic pharyngitis/ tonsillitis (33.99%) |
| Class 2 (n = 3,743): Inflammatory Skin Conditions | 1-Dermatitis and eczema (54.21%)  
2-Dermatitis and eczema (60.65%) |
| Class 3 (n = 1,266): Seizure Disorders and Epilepsy | 1-Seizure Disorder (62.32%)  
2-Seizure Disorder (79.15%)  
3-Seizure Disorder (46.84%) |
| Class 4 (n = 6,446): Asthma | 1-Asthma w/o Status Asthmaticus (50.20%)  
2-Allergic Rhinitis (31.06%)  
2-Asthma w/o Status Asthmaticus (98.37%) |
| Class 5 (n = 28,821): Other, no characteristic morbidity pattern | No diagnoses recorded at ≥ 5% prevalence |
| Class 6 (n = 2,131): Gastrointestinal/Genitourinary Symptoms | 1-Constipation (47.07%)  
2-Constipation (61.76%) |
| Class 7 (n = 1,925): Neurodevelopmental disorders | 1-Autism Spectrum Disorder (26.81%)  
1-Deafness, hearing loss (20.10%)  
2-Autism Spectrum Disorder (47.12%)  
2-Developmental disorder (49.82%)  
3-Autism Spectrum Disorder (29.87%)  
3-Developmental disorder (36.21%) |
| Class 8 (n = 2,925): Physical Symptoms | 1-Fever (30.02%)  
1-Gastroenteritis (22.56%)  
1-Headaches (28.10%)  
1-Nausea, vomiting (20.00%)  
2-Headaches (21.81%) |

Discussion
The preceding study identified patient subgroups with temporal condition patterns that are significantly more common among obese pediatric patients. LCA allows individuals to have a probability of membership in multiple clusters, however our results showed subjects generally had a very high membership probability for a single class (>70%), suggesting shared clinical characterization within the individual groups. Our findings may be used to characterize the prevalence of common conditions among newly obese pediatric patients and to identify pediatric obesity subtypes. The identified subgroups align with prior knowledge on comorbidities associated with childhood obesity, including gastrointestinal, dermatologic, developmental, and sleep disorders, as well as asthma.6,7 Future work will focus on studying the demographic characteristics of patient subgroups and using class membership to predict future health outcomes.

Table 3. Mean Probability of Class Membership (Mean (SD))

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>89.7%</td>
<td>80.64%</td>
<td>86.87%</td>
<td>87.33%</td>
<td>77.77%</td>
<td>81.95%</td>
<td>84.00%</td>
<td>70.21%</td>
</tr>
<tr>
<td></td>
<td>(18.02)</td>
<td>(13.78)</td>
<td>(19.83)</td>
<td>(14.9)</td>
<td>(11.86)</td>
<td>(15.74)</td>
<td>(18.7)</td>
<td>(19.5)</td>
</tr>
</tbody>
</table>

References