Preventive Medicine Reports 10 (2018) 55–61

Contents lists available at ScienceDirect

Preventive Medicine Reports

journal homepage: www.elsevier.com/locate/pmedr

Optimizing community-level surveillance data for pediatric asthma management

Wande O. Benka-Coker[a,*], Sara L. Gale[b], Sylvia J. Brandt[c], John R. Balmes[d,e], Sheryl Magzamen[a]

[a] Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO, USA
[b] Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, USA
[c] Department of Resource Economics, University of Massachusetts, Amherst, MA, USA
[d] Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA, USA
[e] Division of Occupational and Environmental Medicine, University of California, San Francisco, CA, USA

ARTICLE INFO

Keywords:
Asthma
Classification
Risk stratification
Statistical data analysis
Disease management

ABSTRACT

Community-level approaches for pediatric asthma management rely on locally collected information derived primarily from two sources: claims records and school-based surveys. We combined claims and school-based surveillance data, and examined the asthma-related risk patterns among adolescent students.

Symptom data collected from school-based asthma surveys conducted in Oakland, CA were used for case identification and determination of severity levels for students (high and low). Survey data were matched to Medicaid claims data for all asthma-related health care encounters for the year prior to the survey. We then employed recursive partitioning to develop classification trees that identified patterns of demographics and healthcare utilization associated with severity.

A total of 561 students had complete matched data; 86.1% were classified as high-severity, and 13.9% as low-severity asthma. The classification tree consisted of eight subsets: three indicating high severity and five indicating low severity. The risk subsets highlighted varying combinations of non-specific demographic and socioeconomic predictors of asthma prevalence, morbidity and severity. For example, the subset with the highest class-prior probability (92.1%) predicted high-severity asthma and consisted of students without prescribed rescue medication, but with at least one in-clinic nebulizer treatment. The predictive accuracy of the tree-based model was approximately 66.7%, with an estimated 91.1% of high-severity cases and 42.3% of low-severity cases correctly predicted.

Our analysis draws on the strengths of two complementary datasets to provide community-level information on children with asthma, and demonstrates the utility of recursive partitioning methods to explore a combination of features that convey asthma severity.

1. Introduction

Despite recent data showing stabilization in asthma prevalence (Akinbami et al., 2016), childhood asthma morbidity and mortality remain high, particularly in urban communities (Keet et al., 2015). Due to a variety of reasons, including socioeconomic disparities and access to healthcare, asthma diagnosis and assessment of asthma severity are problematic in low-income and non-White populations (Akinbami et al., 2016; Mitchell et al., 2016; Akinbami et al., 2009). The resulting poor asthma control in these groups is characterized largely by increased hospitalizations, emergency department visits and medical costs, health outcomes considered to be avoidable with appropriate management (Vital, 2011; NHLBI, 2007; Barnett and Nurmagambetov, 2011; Gupta et al., 2006).

In conjunction with proper clinical management, community-level surveillance approaches have been suggested as an appropriate strategy to reduce asthma-related morbidity. Community-level data can provide geographically resolved information on asthma prevalence and asthma-related morbidity. These data can add to the general understanding of challenges and solutions for local asthma management, while being detailed enough to decipher unique community patterns of determinants of disease morbidity, and inform asthma control and management efforts (Asher et al., 1995; Busi et al., 2012; Magzamen et al., 2005).

Presently, community-level asthma prevalence data are limited to two commonly described sources: administrative and healthcare claims records, and school-based surveys. Surveillance data based on

* Corresponding author at: Department of Environmental and Radiological Health Sciences, Colorado State University, 1681 Campus Delivery, Fort Collins, CO 80523, USA.
E-mail address: wande.benka-coker@colostate.edu (W.O. Benka-Coker).

https://doi.org/10.1016/j.pmedr.2018.02.004
Received 17 July 2017; Received in revised form 3 January 2018; Accepted 5 February 2018
Available online 08 February 2018

administrative data are generally available and accessible to researchers; these data are particularly useful when the outcomes of interest are objective metrics of healthcare utilization (Morris et al., 1997; Reeves et al., 2006; Roberts et al., 2006; Labrèche et al., 2008; Walsh-Kelly et al., 2008; Dombkowski et al., 2012; Smith et al., 2005), and by extension, important risk factors for asthma severity and management (NHLBI, 2007; Reeves et al., 2006; Roberts et al., 2006). However, claims data may overrepresent the most severe and/or suboptimally managed cases of asthma, rather than reflect the total burden of disease in a community (Piecoro et al., 2001; Dombkowski et al., 2005). Additionally, measures of severity such as a history of symptom type and frequency are largely not found in the claims databases. Results of the objective metrics of asthma severity, such as pulmonary function testing, also tend to be absent from these sources. Consequently, administrative and claims data may not serve as an optimal stand-alone surveillance system for capture of the community burden of asthma. The challenge of use of these data is developing supplementary surveillance measures to augment the highlighted gaps.

School settings represent an alternative ingress point for health surveillance data due to extensive access to members of the target population (Quinn et al., 2006; Redline et al., 2004). School-based surveillance is an efficient way to capture information on asthma-related morbidity in communities with high pediatric asthma prevalence (Bruzzese et al., 2009). School-based surveys may be limited by the self-reported nature of the data and the frequent lack of objective measures of asthma case status (Davis et al., 2008). Further, some of the measures of asthma symptoms and severity vary temporally (Davis et al., 2008). Data (particularly symptom type and frequency) available through surveys can be coupled with claims data to provide a more comprehensive understanding of the landscape of asthma-related morbidity in a community.

We examine the asthma-related healthcare utilization patterns among adolescent students who are clients of a Medicaid managed care program and have completed a school-based asthma questionnaire. This linkage provides a comprehensive dataset with adequate community-level prevalence and severity information in this population of children with asthma. To identify risk factors (demographic and healthcare utilization) that predict asthma severity, we use recursive partitioning analysis to define key pediatric asthma severity subgroups within this population.

2. Methods

2.1. Study population

As part of the CDC-funded Controlling Asthma in American Cities Project, *Oakland Kicks Asthma*™ (OKA) (co-sponsored by the American Lung Association of California; the University of California, Berkeley; Children's Hospital Oakland; and the Oakland Unified School District (OUSD)) conducted school-based asthma surveillance for adolescent students enrolled in OUSD middle schools. From 2003 to 2008, asthma surveillance was conducted in all OUSD middle schools (n = 20) and three high schools at the start of each school year. Methodology and implementation of the asthma surveillance in the OUSD has previously been described (Magzamen et al., 2005). Briefly, a self-administered, 14-question survey based on the International Study of Asthma and Allergy in Childhood (ISAAC) questionnaire, a standardized asthma questionnaire used to describe the prevalence and severity of asthma (Asher et al., 1995), was provided to students during class. The modified ISAAC survey was designed to be short, easy to complete, and provide information not available from routine administrative health forms. Prior to administration of the case-identification survey, parents were sent a letter that described the project; parents were given the option to opt-out of the survey. At the time of survey administration, students were able to decline to participate. All activities conducted under OKA were approved by the OUSD, the Committee for the Protection of Human Subjects at the University of California, Berkeley, and the Institutional Review Board at Colorado State University.

Interpretation of the survey results was based primarily on the National Asthma Education and Prevention Program Expert Panel III Guidelines for the Diagnosis and Treatment of Asthma (NHLBI, 2007). Students who reported a physician diagnosis of asthma as well as a constellation of symptoms associated with asthma-related morbidity were classified as current asthma, and assigned into severity categories (high- and low-severity). The high-severity students reported either a broader range of symptoms, or higher symptom frequency compared to low-severity students; all students who reported an ED visit were classified as high-severity (algorithm available upon request from the authors). All students identified as current asthma were eligible for education and management interventions conducted by OKA. The study survey and classification algorithm are available upon request from authors.

2.2. Survey and medical claims data linkage

OUSD students identified as current asthma were matched to medical claims data provided by the Alameda Alliance for Health (AAH), the not-for-profit county Medicaid umbrella organization that manages and provides health care services for low-income families in Alameda County, CA (Fig. 1). The project established a memorandum of understanding with AAH for data sharing; students were matched by name, date of birth and current address, before de-identifying the data for analysis. Only AAH data for students who had a primary diagnosis code for asthma (ICD-9 code of 493.xx), resided in the city of Oakland, and who had complete records for primary language, race/ethnicity, and residence were used for data analysis.

Healthcare utilization measures were selected based on previously described methods (Brandt et al., 2010). Briefly, billing codes for asthma-relevant health encounters over the year prior to completion of the OKA survey were selected as potential covariates to explain survey-based classification group. Tables A1 and A2 show asthma-related billing codes collected for this analysis, as well as demographic indicators from the AAH database that were included as potential explanatory factors.

For final analysis we only included students with complete covariate set in their records, completed surveys and with no eligibility gaps in health coverage.

2.3. Statistical methods

Asthma severity profiles were created based on the characteristics of individuals in our linked dataset using the Generalized, Unbiased, Interaction Detection and Estimation (GUIDE) tree algorithm (http://www.stat.wisc.edu/~loh/guide.html) (Loh, 2011; Loh, 2009), and asthma severity categories as the outcome. GUIDE is based on recursive partitioning (classification and regression trees), a group of non-parametric, exploratory techniques that capture variation of a single response variable by repeatedly splitting data into homogenous groups based on a set of explanatory variables (Breiman et al., 1984; Strobl et al., 2009; Lemon et al., 2003; De'ath and Fabricius, 2000). Recursive partitioning builds a classification rule to predict class membership (e.g., high- or low-severity) on the basis of its associated covariates (Zhang et al., 2001). Among other advantages over traditional parametric methods (Strobl et al., 2009; Lemon et al., 2003; Afonso et al., 2012; Kuchibhatla and Fillenbaum, 2002; Speybroeck et al., 2004; Kitsantas et al., 2006), recursive partitioning allows for flexibility regarding distributional assumptions and is well suited to data analyses with limited a priori knowledge of variable relationships (Pagán et al., 2009).

Methodology for the GUIDE algorithm is described in detail in Loh (2009). Briefly, GUIDE uses a two-step splitting approach to identify variables that best predict the outcome based on chi-square significance