UMKC
UNIVERSITY OF MISSOURI-KANSAS CITY

The UMKC Center for Economic Information

210 Haag Hall
5100 Rockhill Road
Kansas City, MO 64110
(816) 235-1314
cei.umkc.edu

Center for Economic Information
Working Paper 1602-01
February 17, 2017

Modelling Address Level Asthma Encounters

by
Peter J. Eaton

Abstract

An ordered probit model is estimated for the severity of childhood asthma. Data are from individual childhood asthma health records that are address geocoded to match a survey of exteriorly measured housing condition. Controls for age, gender, race, past asthma encounters, and environmental factors are included in the model. Preliminary results point to a statistically significant impact of exteriorly measured housing structure conditions on the severity of asthma.

Modelling Address Level Asthma Encounters

The literature indicates that there several triggers for asthma (Matsui, et al., 2016). But most statistical work has concentrated on either small samples of individuals participating in field trials, or on analysis of geographically aggregated data. The model we develop is based on two premises: 1) the appropriate scale for analysis of childhood asthma is the individual and 2) the appropriate location for measurement of exposure to asthma triggers is the residential address of the child[1]. The use of the individual as the basic unit of observation avoids the common problem found in the analysis of geographically aggregated data – the modifiable areal unit problem (maup). We examine retrospective health records of individual encounters of children with the Children's Mercy Hospital system. We classify the variables that we hypothesize to systematically have an impact on asthma severity in children into the following categories:

- Demographic factors – age, gender, race/ethnicity
- Environmental factors – exposure to criteria pollutants, proximity to traffic, exposure to rodents
- Climatological variables – weather type, temperature, rainfall, wind speed and direction
- Economic factors – access to health care, socio-economic status, housing conditions

The results that follow are for one year (2001) of the 13 we have available. These results are preliminary. They are the result of an exploratory analysis and have not yet been validated with data from other years. Several specifications were attempted in an attempt to find the appropriate measures to exposure. This is especially true for environmental variables. Other studies are based on much higher levels of geographic aggregation and therefore provide little guidance for this model.

Based on discussions with health professionals, we define a dependent variable that takes on three values indicating increasing severity of asthma associated with the encounter in question. The variable is based on ICD9 codes[2]:

1. Controlled asthma (extrinsic, intrinsic or unspecified – ICD9 codes 493.00, 493.10, 493.90)
2. Asthma with acute exacerbation (extrinsic, intrinsic or unspecified – ICD9 codes 493.02, 493.12, 493.92)
3. Status asthmaticus (extrinsic, intrinsic or unspecified – ICD9 codes 493.01, 493.11, 493.91)

Data

The process of assembling the data for this study is unique, and deserves a detailed description. The main data sources are

- health care encounter records of Children's Mercy Hospital for encounters that occurred in 2001. These records contain the ICD9 code for the encounter, demographic information about the patient (age, gender, race/ethnicity), method of payment
- EPA data from regional monitors was used to measure exposure to criteria pollutants

[1] We recognize that children also spend a significant amount of time at their school. A more general approach would introduce exposure in school buildings as well. Nothing in our model precludes including measurements at multiple locations.

[2] There are also ICD9 codes for chronic obstructive asthma, but these are evidently were not being used in 2001 in the Children's Mercy system..

- NOAA and Spatial Synoptic Classification system[3] for weather
- state departments of transportation for exposure to traffic related pollutants
- housing conditions from the UMKC Center for Economic Information
- the 2000 census for tract level median income and per cent of population below poverty.. , , . These records contain demographic information for the patient, the date of the encounter, ICD9 code(s) for the encounter, and type of health insurance used Mo asthma 11467 – 10275 centerline geocoded

In 2001 there were 15349 asthma related encounters of patients within the Children's Mercy health System network. The first step in the data assembly process was to geocode these encounters to the Center Line File of the Kansas City Metropolitan Area[4]. Over 90% of the encounters (13829) were successfully geocoded to the center line file. The next step was to geocode those 13829 to specific addresses (parcels). The UMKC Center for Economic Information has been conducting curbside surveys of the housing stock since 2000[5]. We matched the housing survey data to the children's health records, for the purpose of testing the curbside survey as a predictor of childhood asthma. This geocoding process established the sample size of 3025 address geocoded asthma encounters. That is, there were 3025 asthma related encounters with the Children's Mercy Health System that had addresses that matched the addresses of residences in the housing conditions survey. With the parcel address of each encounter, we were also able to associate several measures of exposure to EPA criteria pollutants, and proximity of traffic (automobile and railroad) related sources of lung irritants. With the date of the encounter, we are able to associate the weather type for that day. We also counted the number of previous asthma related encounters within the Children's Mercy Health System. The ordered probit model estimated is as follows:

$$ASTHMA\ (1,2,3)_i = \sum_{k=1}^{n_k} \alpha_k X_{ki} + \sum_{j=1}^{n_j} \beta_j Y_{ji} + \sum_{h=1}^{n_h} \gamma_h Z_{hi} + \epsilon_i$$

Where:

$ASTHMA\ (1,2,3)_i$ is an nX1 vector of asthma classifications for encounter i

X_{ki} is an nXk matrix of individual characteristics (age, sex, race/ethnicity, number of prior asthma related encounters) for the patient of encounter i

Y_{ji} is an nXj matrix of observations of housing characteristics associated with encounter I;

Z_{ht} is an nXh matrix of observations of environmental factors associated with the residence of the patient of encounter i

ϵ_i is a random disturbance, assumed to be normally distributed with mean 0 and variance σ^2.

---

[3] http://sheridan.geog.kent.edu/ssc.html

[4] A center line file is a set lines, with each line representing the center of a street segment. Street segments can be conceptualized as a part of a street that in uninterrupted by intersection with another street. An example of one center line street segment would be the 400 block of Main Street, if there is no intersection between 4th Street and 5th Street, and there are intersections at 4th and 5th Streets..

[5] For details, see Neighborhood Housing Conditions Survey, Final Report for the City of Kansas City, Missouri Contract No.08-UMKC01, (UMKC Center for Economic Information, May, 2009) pp. 3-5.

The first stage of our estimation procedure is to use one year of data to calibrate the model. We originally choose 2000 at the calibrating year, but subsequently found in our descriptive data analysis that some of the ICD9 classifications we use to define the dependent variable was only used beginning in October of that year. We therefore switched to 2001 as the calibrating year. Since this type of analysis has not been done before, we are not sure of how to specify certain variables. We experimented with a large number of types of measurement, particularly with regard to the environmental factors. After several steps we arrived at the model specified in Table Y.

Our preliminary findings with regard to environmental variables are

- None of our attempts to include criteria pollutants were successful[6]. That in and of itself is an interesting result. But we do not know if the result stems from the fact that our measurements are not sufficiently accurate, or because the level of these criteria pollutants does not reach a level sufficient to impact the severity of asthma encounters
- proximity to railroads, defined as a residence for a patient that was within a 500 meter buffer of a railroad is weakly associated with more severe asthma encounters. Other proximity measures [to major arterials (150 meter buffer) and highways (500 meter buffer)], were not found to have an association.
- some weather types (based on the Spatial Synoptic Classification system) for the day of the encounter were associated (positively or negatively, depending upon the weather type) with asthma severity. The measure is a vector of 7 dummy variables for eight different weather types) and they performed better than alternatives, such as temperature, rainfall, wind speed and direction. We believe that this classification system is also better than month dummies or other measures of seasonality. The particular weather types with strong positive association with increased asthma severity are dry polar and moist polar. There is a weak positive association with moist moderate weather type. The excluded category for weather type is dry moderate

Individual demographic characteristics were found to be strongly associated with asthma severity. In particular age was strongly and negatively associated. This is to be expected, as asthma becomes more controlled with age. We used dummy variables to combined race and ethnicity into the following categories:

- white, non-latinx
- black, non-latinx
- other, non-latinx
- latinx
- unspecified

White non-latinx was the base. There was non-significant difference between the unspecified category and white. All other categories were more prone to severe encounters. Black, non-latinx was more prone than latinx. The relatively small number of cases for other makes that result dubious.

For each residence of a child in an encounter, 15 different conditions are measured using a 5 point scale that is roughly based on city housing codes, as well as several categorical variables that characterize

[6] Subsequent publications will contain a full accounting of these attempts.

such factors as structure type. The five point scale for each of the 15 characteristics with corresponding values is as follows:

5. Excellent
4. Good
3. Sub-standard
2. Seriously deteriorated
1. Severely deteriorated

We break the 15 characteristics into three general categories – structure characteristics, grounds characteristics, and city infrastructure characteristics. Table X shows the components of each of the general groups.

Table X

Structure	Grounds	Infrastructure
Roof	Private sidewalk/drive	Public sidewalk
Foundation/walls	Lawn/shrubs	Curb
Windows/doors	Nuisance vehicles	Street lighting
Porch	Litter	Catch basin
Exterior Paint	Open storage	Street

For each of the general groups we construct a composite rating that we have found to be useful in other contexts. It is the percent of ratings with a value of 3 or lower (percent substandard or worse). In the model we tried each structure characteristic individually by using a dummy variable for substandard or worse. The best result was for the set of composite variables. The structure characteristic composite variable was weakly associated with severity of asthma. The three variables together had strong association with severity (F value =). Table W contains sample descriptive statistics and Table Z contains the results of the calibrated ordered Probit model.

One noteworthy empirical outcome of this analysis is the fact that it appears that the ICD9 codes do in fact distinguish three ordered outcomes. The “cuts” are not significantly different from the values specified (z score for cut 1 = 1.059, z score for cut 2 = -1.414). Also the model correctly predicts 75.7% of the encounters.

Table W: Sample Descriptive Characteristics	
Male	1809
Female	1209
Mean % Substandard Structure Conditions	14.65%
Mean % Substandard Grounds Conditions	12.84%
Mean % Substandard Infrastructure Conditions	7.32%
White, non-latinx	333
Latinx	168
Black, non-latinx	2435
Other, non-latinx	44
Uspecified race and ethnicity	45
Per Cent within 150 meters of Prinicipal Arterial	31.47%
Per Cent within 500 meters of Highway	24.93%
Per Cent within 500 meters of Railroad	12.13%
Payment Type	
KS Medicaid	82
MO medicaid	2298
Private Insurance	451
Self pay	191
Weather Type	
Dry Moderate	748
Dry Polar	425
Dry Tropical	26
Moist Moderate	379
Moist Polar	353
Moist Tropical	827
Transition	252
NA	15
Mean Prvious Asthma Related Encounters of Patient	3.922

Table Y: Ordered Probit for 2001 (n = 3022)
Standard errors based on Hessian

	Coefficient	*Std. Error*	*z*	*p-value*	
Sub_std structure	0.216232	0.125079	1.7288	0.0839	*
Sub_std grounds	−0.069504	0.151236	−0.4596	0.6458	
Sub_std infra	0.209892	0.166668	1.2593	0.2079	
Age	−0.0187058	0.00520263	−3.5954	0.0003	***
Sex (0=male)	−0.045007	0.0505099	−0.8911	0.3729	
PA_150m	−0.0104851	0.0534782	−0.1961	0.8446	
HWY_500m	−0.0853783	0.0578371	−1.4762	0.1399	
RR_500m	0.109665	0.0746158	1.4697	0.1416	
MO-Medicaid	0.790176	0.137798	5.7343	<0.0001	***
Private_Insurance	0.341286	0.0674693	5.0584	<0.0001	***
Self_pay	0.152769	0.101671	1.5026	0.1329	
Latinx	0.255574	0.132749	1.9252	0.0542	*
Black_non_latinx	0.375334	0.0877029	4.2796	<0.0001	***
Other_non_latinx	0.570563	0.215173	2.6516	0.0080	***
Unspecified	−0.0671077	0.24681	−0.2719	0.7857	
Dry_Polar	0.201431	0.0801391	2.5135	0.0120	**
Dry_Tropical	0.143343	0.255667	0.5607	0.5750	
Moist_Mod	0.141899	0.0843296	1.6827	0.0924	*
Moist_Polar	0.209826	0.0850252	2.4678	0.0136	**
Moist_Tropical	0.0116502	0.0694433	0.1678	0.8668	
Transition	0.0289298	0.0990618	0.2920	0.7703	
Weather_NA	−4.94939	597.254	−0.0083	0.9934	
Prev_encounters	0.0123495	0.00480826	2.5684	0.0102	**
cut1	1.11605	0.109629	10.1802	<0.0001	***
cut2	1.84076	0.112621	16.3448	<0.0001	***

Mean dependent var	1.323958	S.D. dependent var	0.618014
Log-likelihood	−2077.731	Akaike criterion	4205.463
Schwarz criterion	4355.805	Hannan-Quinn	4259.520

Number of cases 'correctly predicted' = 2288 (75.7%)
Likelihood ratio test: Chi-square(23) = 128.377 [0.0000]