A Disaggregate Analysis of the Health Impacts of Transportation System and Urban Design

Mehran Fasihozaman Langerudi
Ph.D. Student
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor Street
Chicago, IL 60607-7023
Tel: (312) 996-0962
Fax: (312) 996-2426
Email: fasihozaman.mehran@gmail.com

Abolfazl (Kouros) Mohammadian, Ph.D.
Associate Professor
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor Street
Chicago, IL 60607-7023
Tel: (312) 996-9840
Fax: (312) 996-2426
Email: kouros@uic.edu

P.S. Sriraj, Ph.D.
Director, Metropolitan Transportation Support Initiative (METSI)
and Research Assistant Professor
Urban Transportation Center
University of Illinois at Chicago
412 S. Peoria St, Suite 340
Chicago, IL 60607
Tel: (312) 413-7568
Email: sriraj@uic.edu

Paper submitted for:
4th Transportation Research Board Conference on
Innovations in Travel Modeling (ITM)
April 30th – May 2nd, 2012 in Tampa, Florida

Abstract

In this paper, a new methodology is used to disaggregate county-level health data into a smaller geography (e.g., census tracts). Then, the disaggregated data is used to estimate various models of individual health condition as a function of socio-demographic, built environment, and transportation system attributes. It is shown that the proposed approach can be applied to disaggregate any aggregate data in an efficient way.

Introduction

The overall built environment, urban form, and elements of the transportation system can shape households' lifestyle; while lifestyle clearly impacts their health and well-being. Although such an interaction between lifestyle and individual health seems highly expected, in many cases the possible magnitude of the impacts are controversial due to the complexity of the involved factors as well as scarcity of reliable data sources. Many researchers have attempted to examine the individual role of various land-use factors on public health (1-11). However, many of such efforts could not explore the health impacts of smaller geographic level regions due to the fact that individuals' health related information is kept strictly confidential and only aggregate level data is publicly available (1). For example the world's largest health data, Behavioral Risk Factor Surveillance System (BRFSS), is only available in County level. That means that BRFSS data records cannot be directly attributed to the local geography, land-use, and built environment variables. It is highly desirable to develop a methodology that can accurately allocate such aggregate data records into smaller geography, so that the causal effects of various health elements and the influence of local physical environment could be properly studied. The primary goal of the current study was to address all these concerns and a methodology is represented to convert aggregate household data into disaggregate format with an acceptable precision process. The methodology is then utilized to examine the effects of land-use and transportation on public health.

Data

The health variables used in this study were extracted from Behavioral Risk Factor Surveillance System (BRFSS) which is tracking health conditions of a large sample of adults in the United States (12). The dataset records demographics, socioeconomics, and health-related information. The dataset is however reported only in county level data as its geographical unit. A portion of the BRFSS 2009 data was extracted that presents records from the six counties of Chicago metropolitan area.

Several other sources of data are utilized in this study including U.S. Census SF3 data (2000) for the disaggregation processing, Census Transportation Planning Package (CTPP 2000), the Census 2000 TIGER/line Geographic Information System data, and National Household Travel Survey (NHTS 2001) data by Federal Highway Administration (FHWA). A large set of transportation and land use variables are obtained by overlaying GIS maps and matching census tract level transportation and land-use shapefiles.

Disaggregation Methodology

Each record of BRFSS dataset represents an adult individual surveyed in the U.S. with a large set of variables including health and demographics information and a county of residence identifier variable. The primary objective of the disaggregation methodology is to allocate each record with a known county of residence into an appropriate census tract. The BRFSS dataset contains both individual and household attributes and demographics such as age, gender, number of household members and household income level, among others. On the other hand, detailed census tract level household and person level data for similar variables could be obtained from Census SF3 files which also provide marginal distributions of those variables. This is the same data that is typically used in many population synthesis practices. "Population synthesis generally utilizes a sample of households at an aggregate geography combined with marginal data on households at a disaggregate geography to generate a set of households which satisfy known marginal at the small area level"(13). Therefore, it seems logical to utilize a similar approach as in a population synthesis to assign BRFSS data records to smaller geography. The only difference would be the use of BRFSS data instead of Census' Public Use Microdata Samples (PUMS) files as the source of disaggregate data sample of households/individuals. The census tract level SF3 data was also applied as the source of marginal distributions. It is noteworthy that the precision of the final disaggregate result decreases as the disaggregate level of geography requested gets smaller and smaller and this was the main reason why the authors kept up with census tract level rather than block group or smaller geographic levels.

Two household level and two person level control variables were used in this study. The number of household members and household income level were considered as the two household level control variables and age and gender variables were used as person level control variables. These four variables are available in both BRFSS and SF3 files with similar definitions. It should be noted that in the new methodology, the objective is just to find an appropriate census tract of residence for the BRFSS records rather than generating a full synthetic population for Illinois. Therefore, the population synthesis procedure has been run over 20 times. Since the procedure is a stochastic process, it is assumed that the higher frequency of observed assign of a person to a census tract yields a better fit of individual's controlled attributes to those of the census tract. Thus, the pattern and frequency of allocation of each sample record to census tracts over 20 runs are carefully examined and individuals are assigned to the census tracts with the highest frequency of observed assignment to that census tract. The procedure was later validated for the purpose of this study.

Methodology Validation

The purpose of the validation procedure was to examine whether the assignment of sample BRFSS individuals to census tracts is valid. As noted earlier due to the stochastic nature of the data synthesize procedure the assignment outcome changes from each iteration to another one. This is mainly due to the nature of the population synthesis which is limited by the control variables that are used in the procedure. In an ideal situation when many control variables are used, the individual can be assigned to the exact census tract home of the person. However, data and computational limitations prevent us from using many control variables. Therefore, to assess the validity of the methodology, the procedure was implemented for five times, that means five different census tract level datasets were generated and compared to examine the validity and consistency of the procedure. Then the five datasets were combined in a regression model to prove that the effect of each set of data is consistent within the regression model. For the purpose of regression analysis, several land use variables of interest were also appended to the data sets. As a result, each dataset consisted of a full scale of variables including individuals' health factors as well as land-use, transportation, and built-environment variables corresponding to the allocated census tract residence of the individual.

A simple regression model was developed using Body Mass Index (BMI) as the dependent variable and several census tract level land use variables as the independent ones for each of the five datasets. For each of the land use variables used in the regression model, five new dummy variables were created in the dataset representing each of the five datasets. The value for each of the dummy variables is the same as its value in the corresponding dataset if the record is attributed to the dataset and 0 otherwise. Table 1, shows few of these land-use dummy variables and the way these five datasets were formed.

Table1. Combined five disaggregate census tract level datasets (partial list of variables)

DN	TU_1	TU_2	TU_3	TU_4	TU_5	RD_1	RD_2	RD_3	RD_4	RD_5	ID_1	ID_2	.	.	.
1	0.13	0	0	0	0	21.8	0	0	0	0	300	0	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.			
.	.	.	.	.	.	.	.	.	.	.	.	.			
.	.	.	.	.	.	.	.	.	.	.	.	.			
2	0	0.35	0	0	0	0	34.2	0	0	0	0	56.8	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.			
.	.	.	.	.	.	.	.	.	.	.	.	.			
.	.	.	.	.	.	.	.	.	.	.	.	.			
3	0	0	0.26	0	0	0	0	16.6	0	0	0	0	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.			
.	.	.	.	.	.	.	.	.	.	.	.	.			
.	.	.	.	.	.	.	.	.	.	.	.	.			
4	0	0	0	0.16	0	0	0	0	25.6	0	0	0	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.			
.	.	.	.	.	.	.	.	.	.	.	.	.			
.	.	.	.	.	.	.	.	.	.	.	.	.			
5	0	0	0	0	0.38	0	0	0	0	15.4	0	0	.	.	

DN: Dataset Number for each record
TU_i: Census Tract Transit Use for records in dataset i and 0 for other records
RD_i: Census Tract Road Density for records in dataset i and 0 for other records
ID_i: Census Tract Intersection Density for records in dataset i and 0 for other records

Equation 1 presents the multiple regression equation that was estimated and Table 2 presents the estimated coefficients of the regression model.

$$BMI = \beta_0 + \beta_1 TU_1 + \beta_2 TU_2 + \beta_3 TU_3 + \beta_4 TU_4 + \beta_5 TU_5 + \alpha_1 ED_1 + \alpha_2 ED_2 + \alpha_3 ED_3 + \alpha_4 ED_4 + \alpha_5 ED_5 + \gamma_1 PD_1 + \gamma_2 PD_2 + \gamma_3 PD_3 + \gamma_4 PD_4 + \gamma_5 PD_5 \quad [1]$$

where:
BMI: Body Mass Index
TU_i: Census Tract Transit Use for dataset i
ED_i: Census Tract Employment Density for dataset i
PD_i : Census Tract Population Density for dataset i

Table 2. Multiple regression model coefficients

Variable	Parameter Estimates	Standard Error	t-value	Pr > |t|
Intercept	0.39964	0.00194	205.52	<.0001
TU_1	0.06190	0.02024	3.06	0.0022
TU_2	0.07966	0.02014	3.96	<.0001
TU_3	0.09086	0.02008	4.53	<.0001
TU_4	0.07033	0.02015	3.49	0.0005
TU_5	0.08912	0.01989	4.48	<.0001
ED_1	-0.00227	0.00055	-4.14	<.0001
ED_2	-0.00220	0.00053	-4.15	<.0001
ED_3	-0.00260	0.00056	-4.61	<.0001
ED_4	-0.00253	0.00057	-4.41	<.0001
ED_5	-0.00316	0.00056	-5.62	<.0001
PD_1	0.00135	0.00032	4.25	<.0001
PD_2	0.00116	0.00030	3.88	<.0001
PD_3	0.00131	0.00032	4.06	<.0001
PD_4	0.00141	0.00032	4.40	<.0001
PD_5	0.00160	0.00032	4.97	<.0001

Statistically, if we show that the estimated coefficients of the same variables (e.g., land-use variable) for the five different datasets are equal with a high degree of significance, it can be proved that the disaggregation process generated consistent results. Equation 2 presents the null hypothesis of this statistical test.

$$H_o: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 \quad [2]$$
$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$$
$$\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$$

The hypothesis of the equally of the coefficients of regression model was tested and it was proved that the null hypothesis could not be rejected at 99% confidence level suggesting that the synthetic procedure of census tract allocation has generated consistent results.

Disaggregate Health Models

As mentioned earlier, the focus of this study is on the disaggregation methodology but the data extracted from the methodology was applied in the health related models. Health indicator variables that were examined in this study include general health, obesity, high blood pressure, high blood cholesterol, and Asthma. Since most of these variables were defined as binary variables in the main dataset, binary choice models were found to be appropriate for the exercise. Table 3, represents several binary probit choice models and their estimated coefficients along with t-statistics of the independent variable in the model with 95% confidence level.

The land-use data that was used in this study was expanded to accompany several mixed land-use information for each census tract. Therefore, density of various land-uses like education centers, retail stores and malls, medical centers, recreation and industrial centers were defined as the number of such centers in a tract divided by the area of the tract. Several interesting results that can be extracted from the table 3 are listed in the following statements:

- High correlation between health factors and demographics could be expected from the coefficients.
- Neo-traditional developments with higher population densities are less consistent with people's general health condition.
- Transit Use has a positive impact on health by lowering obesity, risk of heart attack and asthma rates.
- One of the interesting results of the models is related to several specific land-uses that were examined in the models like positive impact of recreational facilities, malls and retail store on public health.

Table 3. Results of the binary probit model

Variables	General Health	Obesity	Heart Attack	High Blood pressure	High Blood Cholesterol	Asthma
Constant	0.22	-1.37	-2.63	-1.39	-1.15	-0.91
Age	-0.13 (-5.8)	0.035(2.95)	0.03(8.77)	0.03(18.04)	0.03(15.67)	-
Children	-0.08(-1.82)	-	-	-		-
Exercise	0.52(7.97)	-0.18(3.58)	-	-0.145(-2.45)	-0.17(-2.57)	-0.25(-.375)
Education	-	-	-	-	-0.04(-1.8)	-
Income	0.19(13.52)	0.76(2.6)	-0.07(3.31)	-0.07(-5.2)	-0.09(-6.23)	-
Pop Density	-0.004(-2.93)	-	0.006(1.9)	0.0013(2.28)	-	0.002(1.97)
Transit Use	0.07(2.56)	-0.95(-1.9)	-0.64 (-2.45)	-	-	-0.5(-1.86)
Retail Stores	0.35 (1.83)	-	-0.24 (-1.96)	-	-	-
Medical cent	-	-	0.07 (-2.15)	0.16(2.65)	-	-
Recreation Centers	0.002(2.13)	-0.01(-1.85)	-0.004 (-2.83)	-	-0.007(-2.06)	-0.004 (-2.3)

Note: t-statistics are reported in the parentheses.

Conclusions

The study attempts to examine the effects of built-environment, land-use, and transportation on public health. However, due to confidential nature of the individual level health data, such analysis has been typically conducted in an aggregate level of geography. In this paper, a methodology is developed and applied to disaggregate county-level health data into smaller geography of census tract level. Then the disaggregated data is used to estimate various models of individual health condition with specific attention to the effects of local level land-use, built environment, and transportation. The methodology can be utilized in any condition where a survey is conducted in a larger geography unit and there is a need to allocate the record to a smaller geography.

References:

1. Frank, L. D., and P. Engelke, P., (2003). How land use and transportation systems impact public health: A Literature Review of the Relationship Between Physical Activity and Built Form, *ACES: Active community environments initiative working paper 1*, City and Regional Planning Program College of Architecture, Georgia Institute of Technology.
2. Besser, L. and A. L. Dannenburg (2005), Walking to Public Transit: Steps to Help Meet Physical Activity Recommendations, MPH. http://www.cdc.gov/healthyplaces/publications/besser_dannenberg.pdf, accessed July 21, 2011.
3. Ewing, R., T. Schmid, R. Killingsworth, A. Zlot, and S. Raudenbush (2003), Relationship Between Urban Sprawl and Physical Activity, Obesity, and Morbidity, in American Journal of Health Promotion, Vol 18 (1), pp.47-57.
4. Samimi, A., and Kouros Mohammadian (2010). Health Impacts of Urban Development and Transportation Systems. Journal of Urban Planning and Development, 136(3), 208-213
5. Samimi A., Mohammadian A., and Madanizadeh S. (2009). “Effects of Transportation and Built Environment on General Health and Obesity.” Transp. Res. D, 14(1), 67–71.
6. Ewing R., Pendall R., Chen D. (2002) Measuring Sprawl and Its Impact. Washington, DC: Smart Growth America. Available at www.smartgrowthamerica.org/sprawlindex/ sprawlindex.html. Accessed July 21, 2011.
7. Calthorpe, P. (1993). The Next American Metropolis. Princeton Architectural Press 1993, Princeton, New Jersey.
8. Duany A. , E. Plater-Zyberk, Alex Krieger, William R Lennertz, Gund Hall Gallery (1991). Towns and Town-Making Principles. [Cambridge, Mass.] : Harvard University Graduate School of Design ; New York.
9. Boarnet M. and R. Crane (2001). Travel By Design: The Influence of Urban Design on Travel. New York: Oxford University Press
10. Crane R. (1996). On Form Versus Function: Will the New Urbanism Reduce Traffic, or Increase It. Journal of Planning Education and Research, Vol. 15 , pp. 117-126.
11. Kenneth Joh, Marlon G. Boarnet, Mai Thi Nguyen, William Fulton, Wally Siembab and Susan Weaver (2008). Accessibility, Travel Behavior, and New Urbanism: Case Study of Mixed-Use Centers and Auto-Oriented Corridors in the South Bay Region. Transportation Research Board No.2082, 81-89.
12. National Center for Chronic Disease Prevention and Health Promotion, (2009). Behavioral Risk Factor Surveillance System 2009. http://www.cdc.gov/brfss/, accessed 15 October, 2010.
13. Auld, J.A., A. Mohammadian and K. Wies (2009). Population Synthesis with Region-Level Control Variable Aggregation. Journal of Transportation Engineering, 135 (9), 632-639.