



Neighborhood and Environmental Factors Associated with Physical Activity in Prenatal Women.

Montse Fuentes

Statistics Department, NCSU

fuentes@stat.ncsu.edu

In collaboration with B. Reich (Statistics, NCSU)

and Amy Herring (Biostatistics, UNC)

Motivation: Physical activity during pregnancy

- For many health benefits, regular physical activity and exercise are recommended for pregnant women.
- Little is known about the physical activity behavior during pregnancy, particularly on the detailed modes of activity.



- Epidemiologists, policy makers, and city planners are interested in whether characteristics of the physical environment in which women live and work have influence on physical activity levels during pregnancy.



Objectives

The aim of this study was to describe, among a cohort of pregnant women in NC:

- the associations between physical activity and several factors including personal characteristics, meteorological/air quality variables, and neighborhood characteristics.

We have access to unique information about the community environment not captured by the demographic quantities in census data.





Data

- **Subjects:** $n = 1,046$ pregnant women aged 20-39 years residing in four counties in North Carolina where given a physical activity survey at 27-30 weeks' gestation. Women were recruited between January, 2001 and June, 2005.
- **Interview:** Interviewer-administered telephone survey on the mode, frequency and duration, of all physical activity in past seven days during pregnancy.

- **Outcome variables:** weekly MET hours of six types of activity:
 - child care (playing outside with kids, changing diapers...),
 - indoor (house cleaning),
 - outdoor (lawn mowing),
 - recreational (running in the park, working out at the gym),
 - transportation (riding a bike), and
 - work (going up/down stairs in work building).

- **Predictor variables:**

- **Personal characteristics:** income (percent above the poverty level), age, education level, BMI (kg/m^2). Categorical variables to indicate: marital status, race, employment status, gestational diabetes, and whether it is the subject's first birth.
- **Atmospheric variables:** weekly average temperature, precipitation, and ozone.
- **Neighborhood (census block group) variables:** urban/rural area, the speed limit of the subject's street of residence, and three social constructs: *physical incivilities, territoriality, and social spaces*.

Neighborhood characteristics: 3 social constructs

- 17 attributes are used to characterize *neighborhoods* (census block group): condition of yard, presence of no trespassing signs, fences, porches, parks...
- Values were dichotomized by the presence of absence of each attribute for each *street segment* (section of street between 2 intersections) in the neighborhood.
- The proportion of streets within the block group (neighborhood) with the presence of the attributes was calculated. We obtained 17 variables.
- Factor analysis is used to obtain 3 scale weightings from the 17 variables.

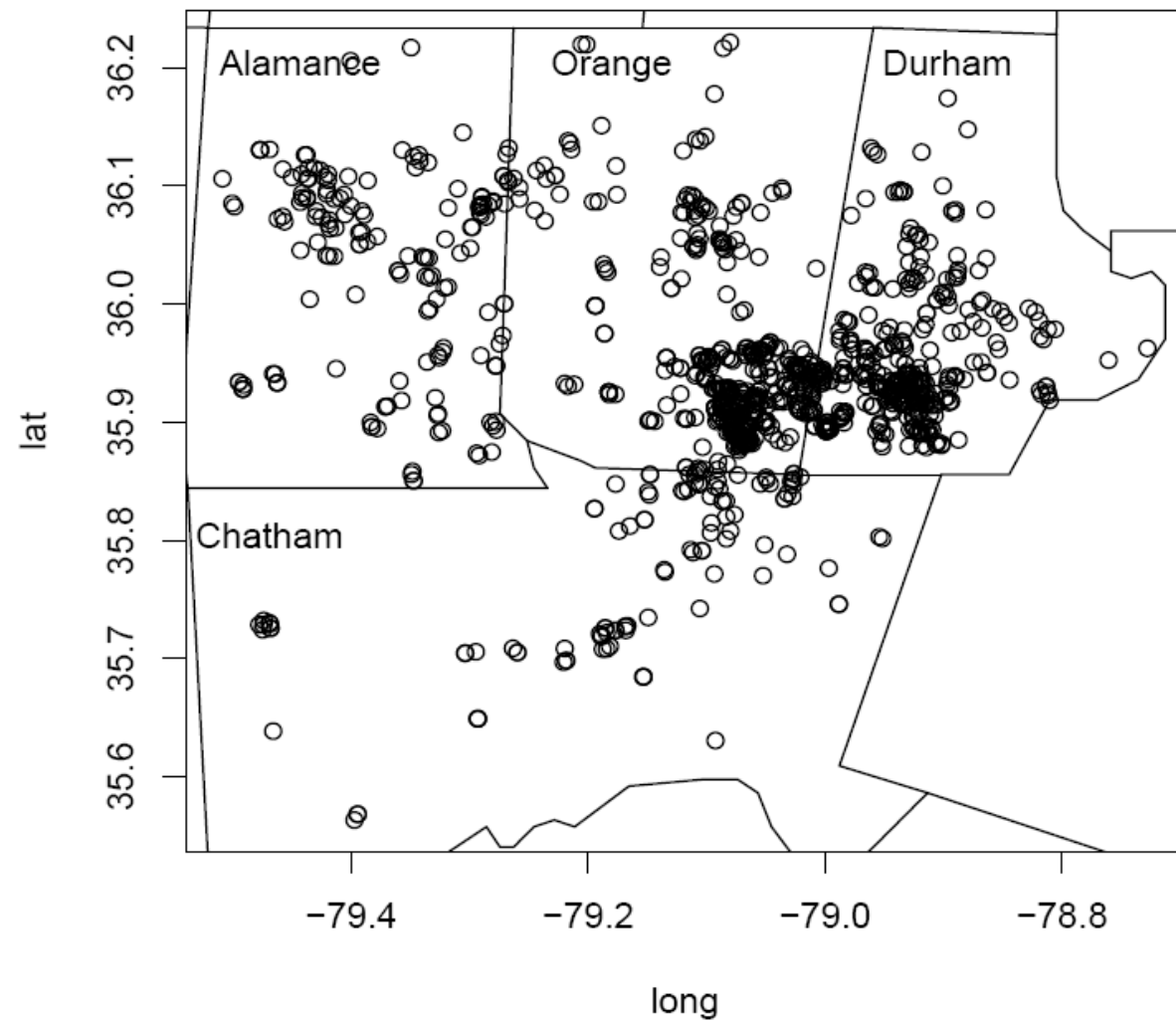
Social constructs (3 scale neighborhood variables).

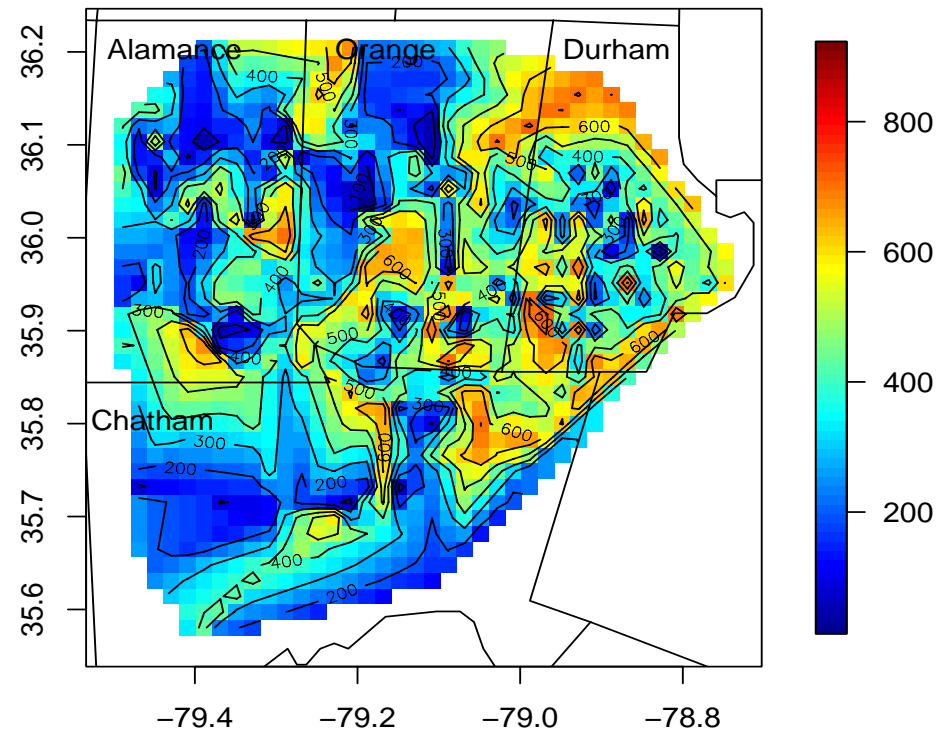
Scale 1 **Physical incivilities:** a combination of physical disorder and poor housing condition. This may communicate decreased local social control and could contribute to crime and further neighborhood deterioration.

Scale 2 **Territoriality:** measures physical and symbolic demarcations of residential property. This may communicate ownership and social control, and may lead to protective effects against crime and adverse community events.

Scale 3 **Social spaces:** measures the presence of people, yards, porches, parks, streets with low speed limits, and sidewalks.

Location of the observations (geocoded addresses of women in cohort)





One of the predictors: Income (percent above the *poverty level*). (*Poverty level in 2008 in U.S. for a family of 4 members is \$21,200 per year.*)

Bayesian Statistics

”The essence of the Bayesian approach is to provide a mathematical rule explaining how you should change your existing beliefs in the light of new evidence. In other words, it **allows scientists to combine new data with their existing knowledge or expertise.**

The canonical example is to imagine that a precocious newborn observes his first sunset, and wonders whether the sun will rise again or not. He assigns equal prior probabilities to both possible outcomes, and represents this by placing one white and one black marble into a bag.

The following day, when the sun rises, the child places another white marble in the bag. The probability that a marble plucked randomly from the bag will be white (ie, the child's degree of belief in future sunrises) has thus gone from a half to two-thirds. After sunrise the next day, the child adds another white marble, and the probability (and thus the degree of belief) goes from two-thirds to three-quarters. And so on.

Gradually, the initial belief that the sun is just as likely as not to rise each morning is modified to become a near-certainty that the sun will always rise.”

From article in the Economist (9/30/00).

Bayesian Statistics

Updating our knowledge with data (Bayesian statistics).

1. We treat *unknowns* (θ) as random variables and we assign them a distribution a priori. **The prior distribution**, $p(\theta)$.
2. We obtain data, y (our observations). We model **the data** given the unknowns (likelihood) $y \sim f_\theta$
3. We update the prior distribution with the data, to obtain a **posterior distribution** for the *unknowns*, which is the distribution of the unknowns given the data (Bayes theorem):

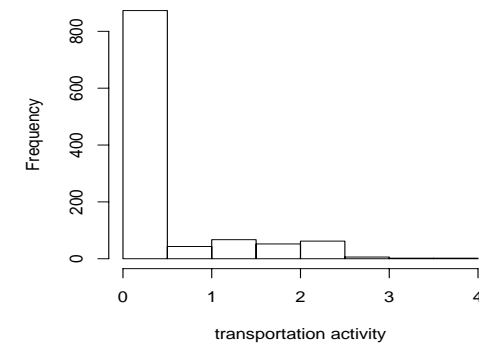
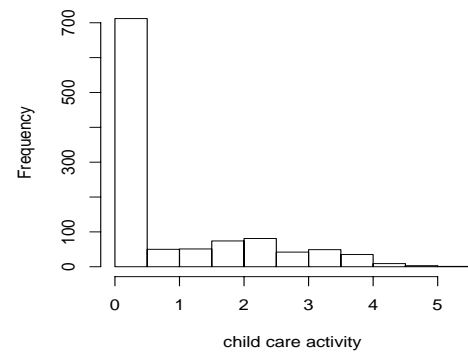
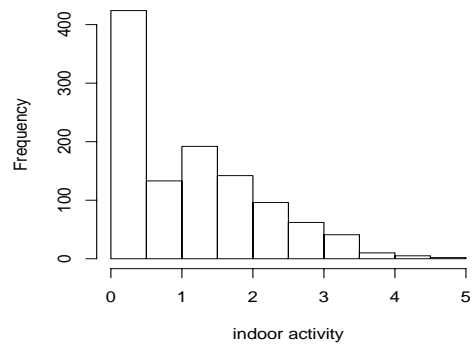
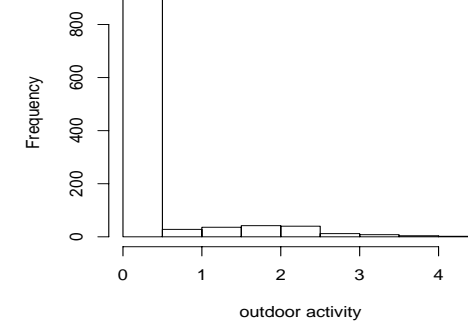
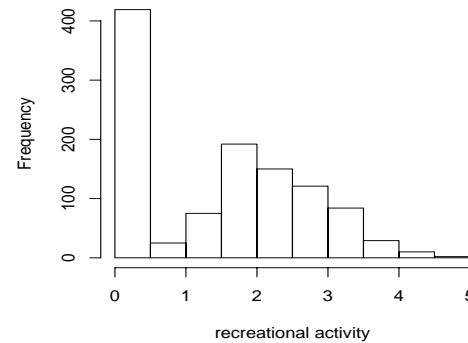
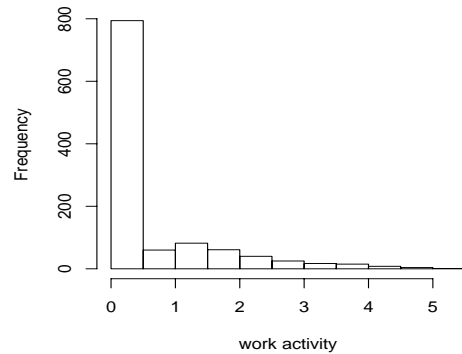
$$p(\theta|y) \propto p(\theta)f_\theta(y)$$

In this work we use a Bayesian approach.

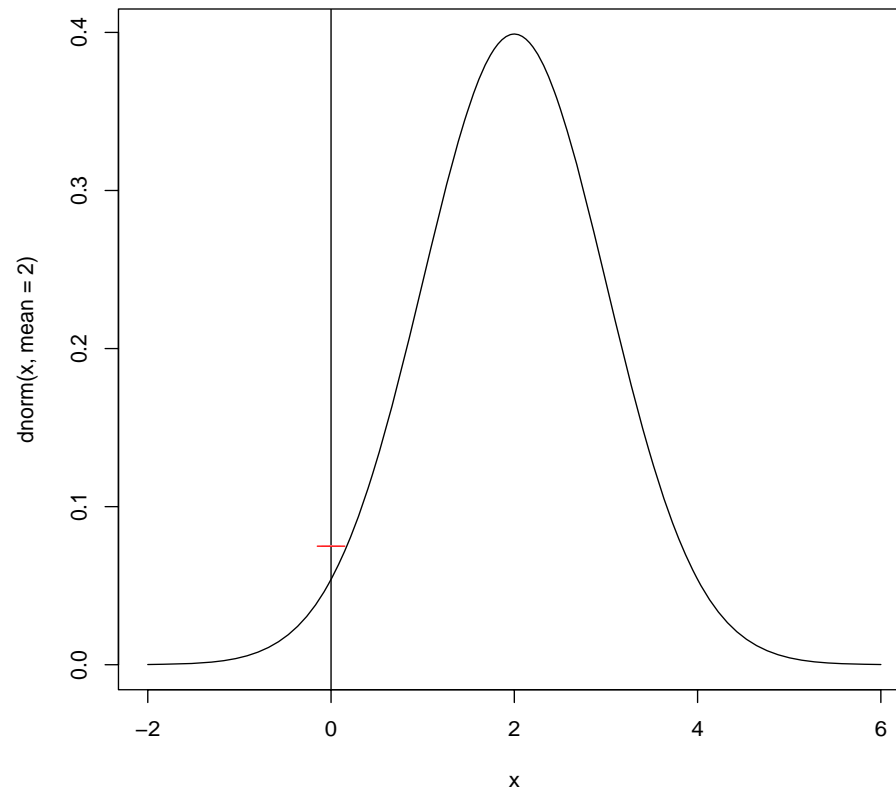
Modeling challenges

- **Multivariate outcomes:** The six activity types are correlated and analyzed simultaneously.
- **Zero-inflation:** Activity is nonnegative and roughly 60% of the observations are zero.
- **Large number of predictors:** We consider 22 possible predictors (all standardized to have mean zero and standard deviation one). They are not all necessary to explain all 6 activity types.
- **Spatially-varying coefficients:** Exploratory analysis suggests that the effect of several of the covariates changes with spatial location.

Zero-inflated distribution for activities



Histograms for for $y = \log(1 + \text{number of MET hours})$.



Truncated normal. Red mark indicates additional mass at zero.

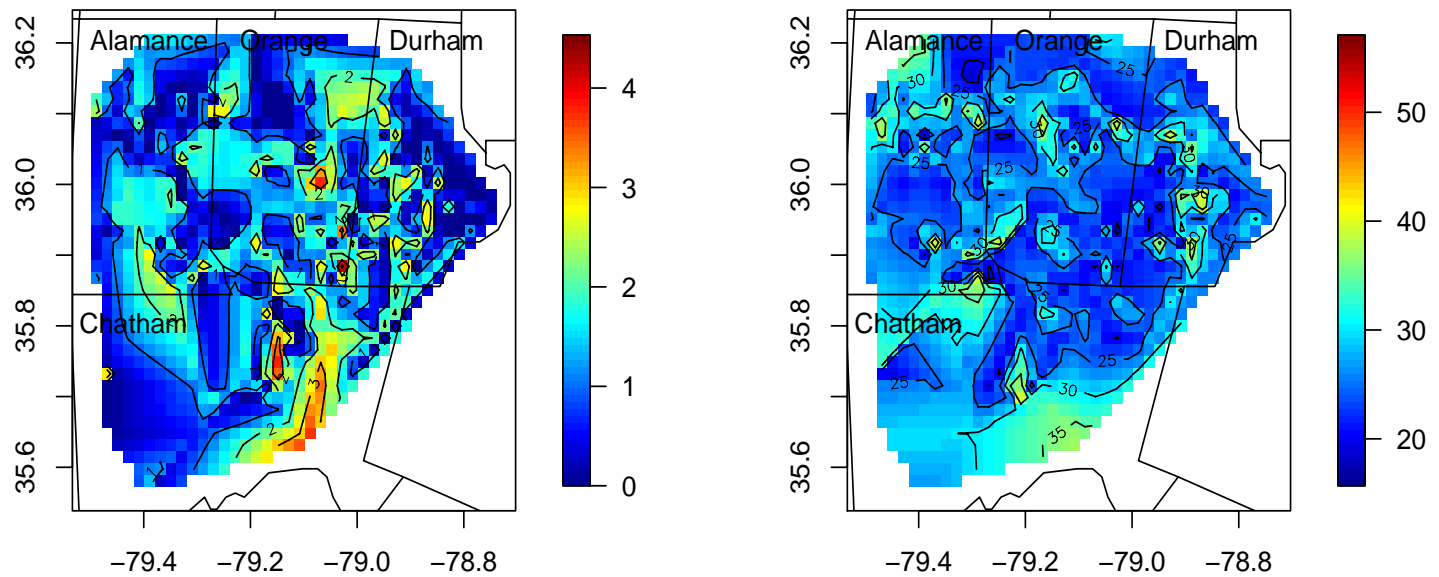
Zero-inflated regression

- Let the activity of type l for the mom at location s_i be

$$y_l(s_i) = \begin{cases} 0 & \text{with probability } p_l \\ z_l(s_i)^+ & \text{with probability } 1 - p_l, \end{cases}$$

- $z_l(s_i)^+ = \max(0, z_l(s_i))$. $z_l(s_i)$ is a latent continuous process
- $(z_1(s_i), \dots, z_6(s_i))' \sim N(\boldsymbol{\mu}(s_i), \Sigma^e)$
- $p_l \sim \text{Unif}(0,1)$ is the additional mass at zero for activity type l that can not be explained by the truncated normal model.
- $\text{Prob}(y_l(s_i) = 0) = p_l + (1 - p_l) * \text{Prob}(z_l(s_i) \leq 0)$, so although p_l is constant across space, $\text{Prob}(y_l(s_i) = 0)$ is not.

Spatial varying covariates and activities



Left: Recreational activity ($\log(1 + \text{met hours})$). Right: BMI km/m^2 .

Spatially-varying regression

- The covariates enter the model through the mean $\mu_l(s_i)$.
- Since $\mu_l(s_i)$ affects $P(y_l(s_i) = 0)$, both zero and non-zero observations inform about $\mu_l(s_i)$.
- The spatially-varying coefficient model of Gelfand et al. allows the regression parameters to be different in different subregions.
- That is $\mu_l(s_i) = \sum_{j=0}^p x_j(s_i)\beta_{lj}(s_i)$ where
 - $x_j(s_i)$ is the value of the j^{th} covariate at site s_i ,
 - $\beta_{lj}(s_i)$ is the corresponding regression parameter for activity type l .

Spatially-varying regression

- We assume $\beta_{lj} = (\beta'_{1j}, \dots, \beta'_{kj})'$ is a Gaussian process ($k = 6$) with
 - $E(\beta_{lj}(s_i)) = \mu_{lj}$
 - $\text{Cov}(\beta_{lj}(s_i), \beta_{l'j'}(s'_i)) = K_j(s_i, s'_i)\Sigma_j(l, l')$.
- μ_{lj} is the spatial average effect
- $\Sigma_j = \{\Sigma_j(l, l')\}$ ($k \times k$) controls the correlation across activity types.
- $K_j(s_i, s'_i)$ is the spatial correlation function ($n \times n$). We assume K_j is a standard parametric spatial correlation function (Matérn).

Variable selection

- We consider 22 covariates. However, including all 22 covariates in the spatially-varying coefficients model gives an overparameterized model.
- Not all of the covariates are needed for every activity type and the effect of most covariates do not vary spatially.

- We propose a stochastic search variable selection model that allows the x_j to enter the model in one of three ways:
 1. x_j is **removed** from the model (variable j no effect on any outcome) and $\beta_{lj}(s_i) = 0$ for all s_i .
 2. x_j 's effect on each outcome is **constant across space** and $\beta_{lj}(s_i) = \mu_{lj}$ for all s_i , or
 3. x_j 's effect **varies spatially** and $\beta_{lj}(s_i)$ varies from location-to-location.

Variable selection - average effects

The overall average effect μ_{lj} is given the spike and slab prior, popularized for variable selection in the usual linear regression setting by George and McCullough.

- We assume $\mu_{lj} = \gamma_{1lj}\alpha_{lj}$ where
- $\gamma_{1lj} \in \{0, 1\}$
- $\alpha_{lj} \sim N(0, \sigma_j)$
- If $\gamma_{1lj} = 0$ then $\mu_{lj} = 0$ and x_j 's average (over space) effect on activity type l is zero.

Variable selection - spatially-varying effects

To determine which regression parameters should be allowed to vary spatially, the prior for the columns and rows of the covariance matrix Σ_j ($k \times k$) (for the activity types) also have positive mass at zero.

- We assume $\Sigma_j = \text{diag}(\gamma_{21j}, \dots, \gamma_{2kj}) \Omega_j \text{diag}(\gamma_{21j}, \dots, \gamma_{2kj})$ where
- $\gamma_{2lj} \in \{0, 1\}$
- $\Omega_j \sim \text{InvWishart}(k+1+.1 \text{ d.f.}, I_k \text{ scale matrix})$
- If $\gamma_{2lj} = 0$ then $\beta_{lj} = \mu_{lj}$ for all s_i and the effect is x_j is constant across space for activity type l .

Priors for inclusion indicators

- The prior for the binary variable inclusion indicators ensures that the effects are not allowed to vary spatially unless the overall average is non-zero.

$$p(\gamma_{1lj}, \gamma_{2lj}) = \begin{cases} 1 - \pi_{1j}, & \gamma_{1lj} = 0 \text{ and } \gamma_{2lj} = 0 \\ \pi_{1j}(1 - \pi_{2j}), & \gamma_{1lj} = 1 \text{ and } \gamma_{2lj} = 0 \\ \pi_{1j}\pi_{2j}, & \gamma_{1lj} = 1 \text{ and } \gamma_{2lj} = 1 \end{cases}$$

- π_{1j} is the prior probability that x_j is included in the model (either constant or spatially-varying)
- π_{2j} is the conditional probability that the effect of x_j varies spatially given that it is included in the model for activity l .

Hyperpriors for inclusion indicators

- Bayesian variable selection can be sensitive to the choice of π_{1j} and π_{2j} and variance σ_j^2 .
- In this multivariate variable selection setting we can alleviate some of these problems by pooling information across activities.
- We assume π_{1j} , π_{2j} , and σ_j^2 are shared across activities and
 - $\pi_{1j} \sim U(0,1)$
 - $\pi_{2j} \sim U(0,1)$
 - $\sigma_j^2 \sim \text{InvGamma}(0.5,0.5)$.
- This hierarchical model facilitates pooling information regarding the importance of x_j across activity types.

Results - model comparisons via DIC

- We compare models using DIC (Speigelhalter et al., 2002)
- $DIC = \bar{D} + p_D$ where
- \bar{D} , the posterior mean of the deviance, measures fit.
- p_D , the effective number of parameters, measures complexity.
- The full model has smaller DIC than several simpler models.

Model	\bar{D}	p_D	DIC
Full model	14605.2	170.5	14775.7
No hierarchical model for π_{1j} , π_{2j} and σ_j^2	14605.4	190.1	14795.5
No spatially-varying coefficients	14916.0	60.3	14855.6
No correlation across activities	14731.7	159.3	14891.0

Full model - inclusion probabilities ($E(\pi_{1jl}|y)$)

Gray (black) indicates the inclusion prob is greater than 0.5 (0.9); “+” and “-” is the sign of $E(\mu_{jl}|y)$.

	Work	Rec	Out	In	Care	Trans
Income		+				
Education	-	+	+	-	+	+
Black	+	-	-		-	+
Married						
Age	-	+	+	+		
First birth	+	+	-		-	+
BMI		-	-			
Weight gain						
Gest diabetes	+					
Hyper/Pih/Pre ecl						
Smoker	+					
General health						
Working	+	-	+	-	-	+
Ozone – high edu						
Ozone – low edu						
Temperature		+	+	-		
Precipitation						
Inciviliaties						
Territoriality						
Social spaces						
Urban	+	-	-	-	+	+
Speed limit	+	-	+	+	+	+

Results

- The most important predictors of physical activity are: education, race, first birth, employment status, urban, speed limit of the street of residence.
- Effect of urban: mothers in urban areas engage in more work, child care, and transportation activity and less indoor, outdoor, and recreational activity.
- Effect of race: black mothers engage in more work and transportation activity and less recreational, outdoor, and child care activity.

- Compared with the simpler model with $\pi_{1j} = \pi_{2j} = 0.5$ (next slide), the hierarchical model for the prior inclusion probabilities π_{1j} and π_{2j} results in more consistent variable selection across activities.

Simple model - inclusion probabilities ($E(\pi_{1jl}|y)$)

Gray (black) indicates the inclusion prob is greater than 0.5 (0.9); “+” and “-” is the sign of $E(\mu_{jl}|y)$.

	Work	Rec	Out	In	Care	Trans
Income		+				-
Education	+	+	+	-	+	+
Black	+	-	-		-	
Married						
Age		+	+			
First birth	+	+			-	+
BMI		-				
Weight gain				+		
Gest diabetes	+					
Hyper/Pih/Pre ecl						
Smoker	+					
General health		+				
Working	+	-				+
Ozone – high edu					-	
Ozone – low edu			+			
Temperature		+	+			
Precipitation						
Inciviliaties						
Territoriality						
Social spaces						
Urban	+	-	-	-		+
Speed limit					+	

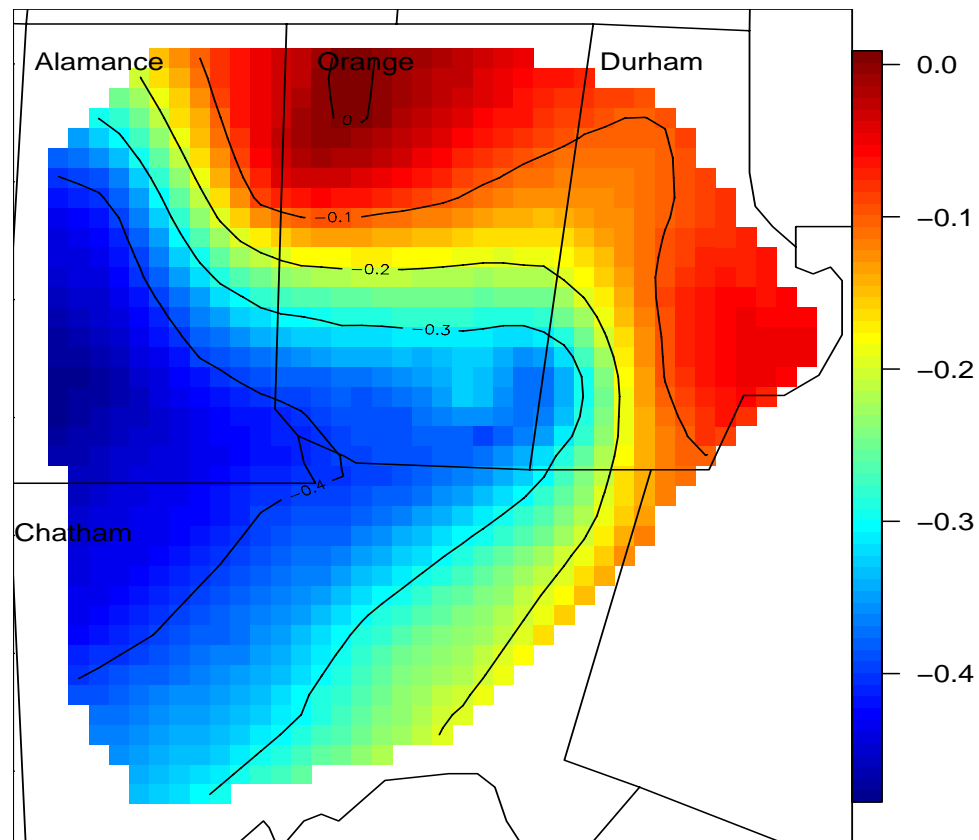
The probability of having spatial -varying effects ($E(\pi_{2jl}|y)$)

Gray (black) indicates the prob is greater than 0.5 (0.9)

	Work	Rec	Out	In	Care	Trans
Income						
Education						
Black						
Married						
Age						
First birth						
BMI						
Weight Gain						
Gest diabetes						
Hyper/Pih/Pre ecl						
Smoker						
General health						
Working						
Ozone – high edu						
Ozone – low edu						
Temperature						
Precipitation						
Inciviliaties						
Territoriality						
Social spaces						
Urban						
Speed limit						

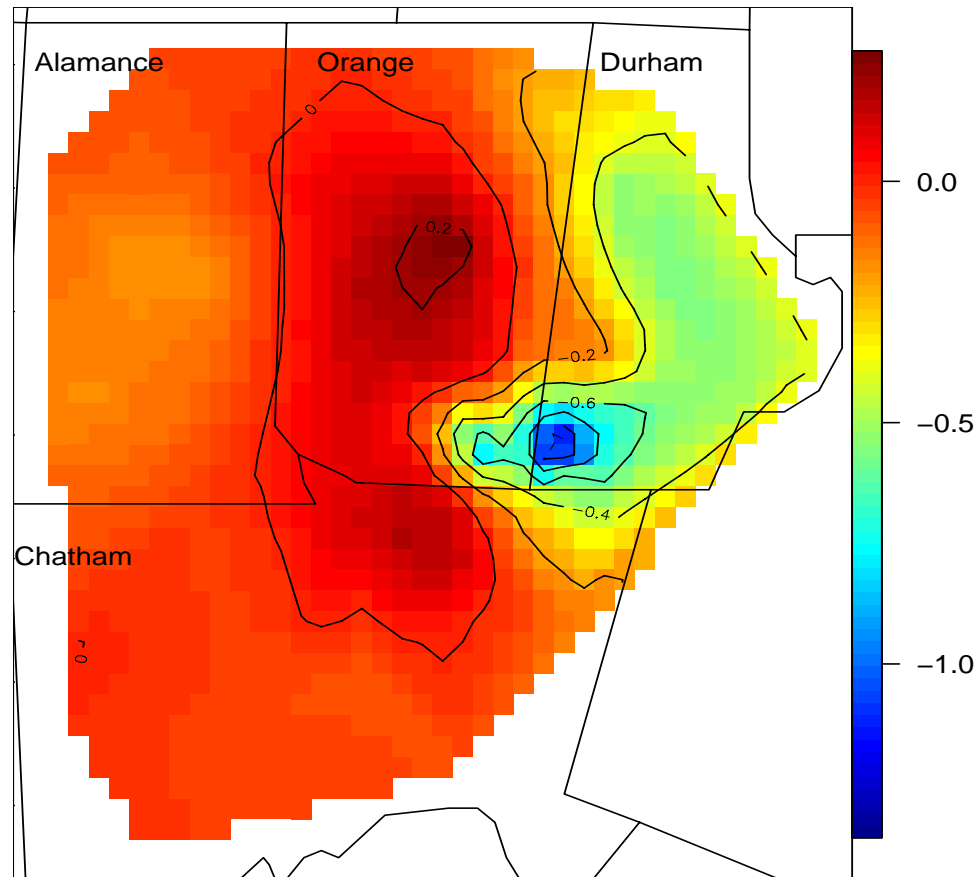
Spatially-varying effect of BMI on recreational activity

BMI has more effect in rural Chatham and Alamance counties.



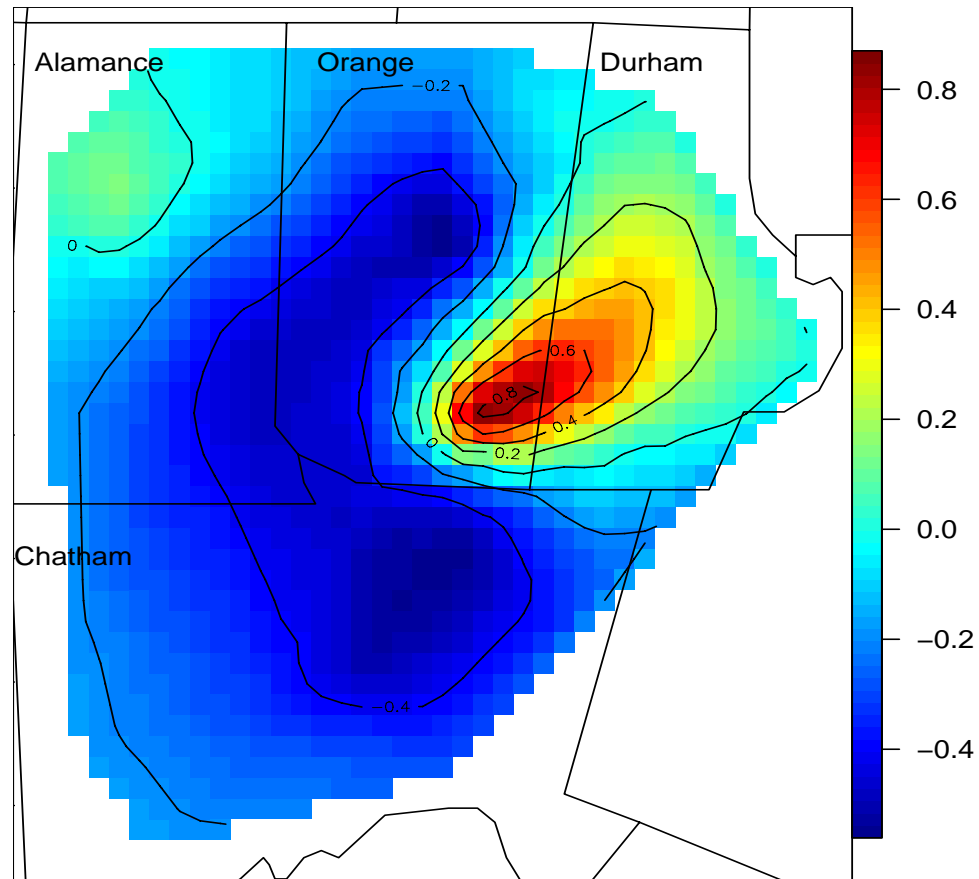
Effect of speed limit on recreational activity

Speed limit has more effect in Chapel Hill/Research Triangle Park.



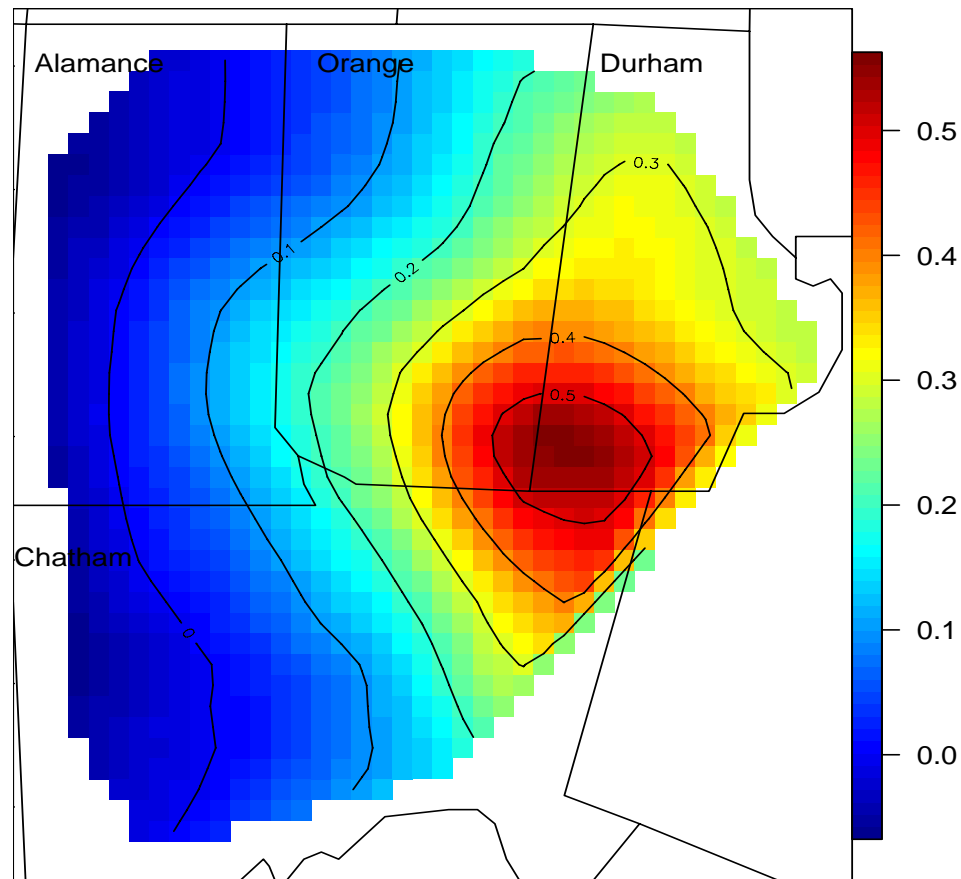
Effect of income on recreational activity

Income has more effect in Chapel Hill (easy access to parks, gyms..).



Effect of Education on recreational activity

Education has more effect in Research Triangle Park.



Summary of results

- Effects vary across space:
 - For recreational activities, BMI has a significant *negative* effect that is space dependent (less effect on wealthier areas). In those areas people have more easy access to recreational activities (gym, parks...).
 - For recreational activities, speed limit matters more in areas with more speed limit variability.
 - For recreational activities: Income has higher effect in the Chapel Hill area (good neighborhoods, easy access to parks, gyms..).
 - For recreational activities: Education has higher effect in the RTP area (easy access to trails, most likely women work in

RTP with easy access recreational activities during lunch
(trails, gyms...).

- Education is important:
 - Women are more active when they have higher education.
 - In the highly educated group (\geq high school), ozone has a negative effect on child care activities (playing with kids outside...).
 - In the lower educated group, there is more outdoor activity on higher ozone days (this effect is not significant in the highly educated group).

- There is more recreational activity if primiparous (less distracted by other kids). Less care activity if primiparous (no other kids).
- More work activity if the woman is working, less time for rec activity if she is working. More transportation activity if she is working (she has to get to work).
- There is a space-dependent effect, and neighborhood characteristics play an important role. However, the 3 social constructs do not have an effect on any of the physical activity types.

Public health interventions, urban planning and public policy should consider address barriers and improve facilitators to increase physical activity within one's neighborhood, plausibly improving the health of our pregnant populations.