WHERE THERE'S SMOKE: CIGARETTE USE, SOCIAL ACCEPTABILITY, AND SPATIAL APPROACHES TO MULTILEVEL MODELING

Heather A. O'Connell

Postdoctoral Fellow

Kinder Institute for Urban Research, Rice University

ABSTRACT

I contribute to the longstanding body of literature on how context is related to health and other individual outcomes by assessing the added value of combining multilevel and spatial modeling techniques. This methodological approach leads to substantive contributions to the smoking literature, including improved clarity on the central contextual factors and the examination of one manifestation of the social acceptability hypothesis. Both contributions help provide a clearer picture of the associations between county-level characteristics and the individual-level odds of smoking during pregnancy. For this analysis I use restricted-use natality data from the Vital Statistics, and county-level data from the 2005-9 ACS. The results suggest that spatial modeling is still critical even in a multilevel framework. In addition, I argue that processes related to social acceptability and spatial diffusion underlie the relationships linking racial/ethnic minority concentration to lower overall odds of smoking.

WHERE THERE'S SMOKE: CIGARETTE USE, SOCIAL ACCEPTABILITY, AND SPATIAL APPROACHES TO MULTILEVEL MODELING

Place and space are integral to understanding how social processes unfold. The call for more careful attention to how we treat place and space has been expressed within the health literature (Cummins et al. 2007; Stafford, Duke-Williams, and Shelton 2008; also see Logan 2012), and scholars have begun to incorporate the two concepts simultaneously within empirical research (see Chaix, Merio, and Chauvin 2004; Crowder and South 2008, 2011; Morenoff 2003; Mu et al. 2015; Perchoux et al. 2014; Savitz and Raudenbush 2009; Xu, Logan, and Short 2014), yet the implementation of these advancements is isolated. Focusing on one component of the overlap between place and space, I aim to promote the use of spatially-informed multilevel models by drawing attention to the associated conceptual and methodological benefits. In addition, I extend previous research that combines hierarchical with spatial modeling by demonstrating how to obtain Moran's *I* estimates and spatial model diagnostics – necessary, yet unavailable (in the multilevel context), guiding tools. These and related methodological advancements discussed below provide "greater focus on the position of places relative to each other" and the substantive implications of spatial autocorrelation (Cummins et al. 2007:1832) and are thus central to our pursuit of a more sophisticated and realistic understanding of how place and space matter for health.

A long line of research has demonstrated the theoretical importance of accounting for hierarchical and spatial structures. Multilevel modeling – encompassing both linear and generalized modeling approaches – can be used to study how place relates to individual-level outcomes, and statistically adjusts for the clustering of individuals within the same place (Bryk and Ruadenbush 1992). Using a similar logic, spatial models address theoretical questions and statistical issues related to how the spatial position of a place in relation to other places affects social processes and statistical estimates (Cliff and Ord 1973, 1981). When we do not account for the clustering of individual observations and the relative spatial position of places, we have a higher risk of coming to the wrong conclusion than otherwise expected (Bryk and Ruadenbush 1992; Cliff and Ord 1973, 1981). Despite the interconnected issues that multilevel and spatial modeling techniques address, research has yet to address both aspects simultaneously (however, see especially Xu et al. 2014). I argue that combining them is necessary for advancing our understanding of how context is related to health and other social phenomena because ignoring space could introduce bias into our estimates of contextual associations as well as leave holes in our theoretical models.

I demonstrate the necessity of spatial approaches to multilevel modeling by examining the relationship between local context and smoking among pregnant women in the United States. Using maternal smoking to illustrate this extension is ideal because recent research suggests strong connections between county context and the odds of smoking during pregnancy (Shoff and Yang 2013). In addition, the role of social acceptability in smoking behaviors presents a superlative opportunity to assess spatial processes in health because this social process is expected to have spatial manifestations (for research discussing social acceptability and smoking see e.g., Alesci, Forster, and Blaine 2003; Botvin et al. 1992).

Through this project, I argue that contextual relationships – however defined – cannot be properly assessed without considering concepts related to space. I take one step towards addressing the interlocking nature of place and space by combining multilevel and spatial regression analysis techniques. The application of this methodological extension suggests that previous work using just multilevel modeling may have overestimated contextual associations. In addition, the results extend the scope of previous contextual analyses of maternal smoking by providing support for social acceptability models of smoking behavior, and by more accurately identifying contextual factors relevant for explaining smoking. In what follows I discuss previous research on smoking, as well as work on spatial approaches to multilevel modeling. I then develop the methodological extensions before providing a demonstration of this approach using individual natality data that is linked to county characteristics.

THE ROLE OF CONTEXT IN MATERNAL SMOKING: SPACE FOR ADVANCEMENT

Increasing focus on how the surrounding environment affects individuals has led scholars to examine the impact of local factors on health outcomes and behaviors (see e.g., Boardman 2004; Kimbro 2009; Kimbro and Denney 2013; Yang and Matthews 2010; Morenoff 2003; Taylor, Repetti, and Seeman 1997; Yang, Matthews, and Shoff 2011). The primary contextual conditions examined are those associated with economic advantage and disadvantage (see e.g., Brooks-Gunn, Duncan, and Aber 1997; Clarke et al. 2014). The role of local socioeconomic status (SES) in shaping the health of all residents is an important connection to address – it is consistently influential across outcomes, and may shape health through a myriad of pathways.

However, scholars have brought attention to additional important factors (e.g., rurality, social capital/cohesion, etc.; see Patterson et al. 2004; Shoff and Yang 2013). Of most relevance to the present research is the significance and interpretation of racial composition variables. Previous research repeatedly shows that the racial composition of a county is related to smoking (Shaw, Pickett, and Wilkinson 2010; Shaw and Pickett 2013; Shoff and Yang 2013; for research at the census tract level see Nkansah-Amankra 2010). Counterintuitively, at least at face-value, places with higher concentrations of Hispanics and non-Hispanic blacks have *lower* average individual-level odds of smoking during pregnancy.

But to what extent do any of these contextual relationships remain net of spatial processes? The substantive significance of previous findings rests on the assumption that the associations are unaffected by spatial processes, which is perhaps a tenuous assumption. Indeed, research suggests caution when interpreting results that do not assess their residuals for spatial autocorrelation (Voss et al. 2006). Therefore, I extend previous research on the contextual factors associated with maternal smoking by assessing the robustness of previous results to spatial processes. In addition, I highlight a difficult to capture process that is of potential significance – social acceptability.

THE ELUSIVE LINK BETWEEN SOCIAL ACCEPTABILITY AND SMOKING

Attention to the topic of social acceptability has entered into a range of health research. Of most relevance, scholars have discussed it within the context of multiple dimensions of smoking (e.g., Afifi et al. 2013; Albers et al. 2004; Alesci et al. 2003; Botvin et al. 1992; Daly et al. 1993; Landman, Cortese and Glantz 2008; Thomson et al. 2005). Research suggests that the smoking behavior of young women is strongly related to the smoking behavior of peers and how acceptable peers find smoking to be (Daly et al. 1993). In addition, ideas related to acceptability have been linked to weight outcomes via dietary and exercising norms (see e.g., Ajilore et al. 2014; Hruscha et al. 2011). Despite evidence of its import, we have a limited ability to account for this process without specialized survey questions.

The consequences of this methodological restriction are particularly evident within the smoking literature that aims to assess the relevance of contextual factors. As mentioned above, research has identified a link between an individual's odds of smoking and the racial composition of their local context (Nkansah-Amankra 2010; Shaw et al. 2010; Shaw and Pickett 2013; Shoff and Yang 2013). Since this association is net of individual-level characteristics, including reported race, it cannot be explained using compositional arguments. Instead, some have suggested that it is related to health behavior norms (Shaw and Pickett 2013). It is possible that concentrations of populations with lower odds of maternal smoking during pregnancy, including but not limited to blacks and Hispanics, reduces the odds that others will engage in the behavior because it is seen as socially unacceptable. However, despite such theorizing research has not been able to assess this pathway. I argue that addressing spatial processes, namely those related to spatial contagion or diffusion, offer one – admittedly indirect – means of assessing the role of social acceptability in this and other contextual relationships. And by isolating social acceptability processes manifested spatially, I will be better able to identify the contextual factors most relevant to shaping smoking.

In developing a spatially-based social acceptability argument, I suggest that the average probability of a pregnant woman smoking in neighboring counties will be positively related to the probability that a pregnant woman will smoke in the focal county. Although I cannot test the mechanism in this paper, the theoretical implication is that this positive spatial relationship is due to the perceived social acceptability of smoking while pregnant that is tied to the frequency of an act in an area. In addition, I note that the link between frequency and acceptability is not a simple one as the forces are likely reinforcing. However, using frequency as an approximation is appropriate for at least a baseline estimate of this process, and will provide a general accounting of acceptability's possible influence on previously established explanations for maternal smoking (e.g., local SES and racial composition). Of primary importance here is the ability to purge contextual factors of the spatial manifestation of this social process. Detailing the specifics of acceptability processes will be a central challenge for future research on health. But now I turn to a discussion of the multilevel and spatial approaches used to address the relationships laid out above.

INCORPORATING SPACE INTO MULTILEVEL MODELS: METHODOLOGICAL AND CONCEPTUAL CONSIDERATIONS

Multilevel modeling approaches have become increasingly popular in the most recent decade. However, in our excitement to take advantage of the benefits of hierarchical linear modeling and its generalized forms (hence forth referenced simply as "HLM"), researchers have all too often forgotten that HLM is prey to the same statistical concerns as are standard regression analyses. This includes, but is not limited to, concerns regarding the spatial independence of our residuals when analyzing geographically contiguous units (e.g., census tracts, school districts, counties). As proponents of HLM have argued, ensuring proper modeling is necessary for drawing well-informed conclusions from our results (see e.g., Teachman and Crowder 2002; Bryk and Ruadenbush 1992). I aim to take this goal a step further by adding spatial considerations to the HLM approach, which I argue is necessary for the accurate assessment of significant contextual processes.

Before further discussing their benefits, a brief review of spatial methods is in order. To begin, Moran's *I* statistics are the primary means of quantifying spatial autocorrelation, and subsequently identifying when models that account for the spatial structure of the data are necessary. Traditionally, any significant level of autocorrelation has motivated using a spatial modelling approach, but Moran's *I* values above 0.10 are of most concern. After identifying spatial autocorrelation among model residuals researchers use diagnostic tools to determine the most appropriate spatial model – either the spatial lag or spatial error. Both approaches to spatial modeling are consistent with a statistical link among neighbors, yet they have distinct theoretical implications. Spatial lag models estimate how the average of neighboring values of the dependent variable relate to the value of the dependent variable in a focal county. The guiding theoretical explanation for this type of association is spatial diffusion, or contagion, and is often linked to social processes like the sharing of ideas. In contrast to a lag, the purpose of a spatial error model is to purge the data of unmeasured spatial processes that result in the correlation of the dependent variable, and subsequently the residuals, across places. Essentially, a spatial error model treats spatial dependence (i.e., the correlation between neighboring places' residuals) as a nuisance rather than as a result of substantive processes related to diffusion. Given their distinct assumptions, distinguishing between these two statistical approaches is paramount to guiding future theoretical development as using a spatial lag instead of a spatial error model could suggest contagion processes despite the absence of such a process.

Spatial considerations have been incorporated specifically into multilevel research in a variety of ways, but are only in their infancy. On the methodological side, Savitz and Raudenbush (2009) have used a spatial approach to HLM to improve the measurement of neighborhood variables (similarly see Chaix et al. 2004; Mu et al. 2015). In addition, the most recent HLM 7.0 software (Raudenbush, Bryk, and Congdon 2011) now includes the option to estimate a spatial lag hierarchical model. However, only a handful of studies have combined spatial processes with multilevel modeling to address substantive, social questions (Crowder and

South 2008, 2011; Morenoff 2003; Xu et al. 2014). Morenoff (2003) extends our understanding of the relationship between neighborhood disadvantage and low birth weight outcomes by incorporating a spatially lagged measure of crime in the second-level of his HLM analysis. Whether referencing an independent or dependent variable, a spatial lag reflects the average value of a variable for the geographic units surrounding a given unit. Extending the modeling options discussed above, using a spatially lagged independent variable assesses the role of extra-local processes, and can also provide a more realistic estimate of factors related to the local context (also see Crowder and South 2008, 2011). While this type of theoretical extension is of great interest, it does not necessarily address lingering statistical concerns regarding spatial autocorrelation among the level-2 residuals in a multilevel model. More recent research has picked up on the latter issue by adding spatially correlated random effects to a standard multilevel Poisson model (Xu et al. 2014). As the authors argue, this approach accounts for the spatial dependence structure of the data, and is therefore analogous to a spatial error model. Xu et al. (2014) demonstrate significant statistical and substantive improvements of the spatial model over the standard, aspatial multilevel model.

Despite these advancements and the conceptual benefits of employing spatial techniques, widespread use of spatially-informed approaches to HLM has yet to transpire. I aim to spur on this movement through a fresh analysis that combines multilevel modeling with spatial modeling approaches. Recent research making similar methodological contributions offers a sophisticated discussion of a myriad of ways to bring space into conversation with contextual analyses (Xu et al. 2014). My work provides an extension to this research in two ways: first, I provide new information on how to incorporate supplementary aspects of spatial regression analysis techniques when using the user-friendly HLM 7.0 software (e.g., estimating Moran's *I* statistics and spatial modeling diagnostics); and second, I make the spatial HLM modeling process more transparent, and therefore accessible, by providing a detailed discussion of my approach in order to promote the use of this modeling strategy (when appropriate).

Focusing on the implications for health research, I suggest two primary benefits of these spatial tools to contextual analyses. Foremost, adjusting for spatial autocorrelation among residuals is necessary for purging coefficient estimates of residual bias, and thus is central to maintaining confidence in our results. Any type of correlation among the residuals – be it due to an omitted variable or, in this case, spatial proximity – puts us at additional risk of concluding that an association is statistically significant when the relationship is not socially relevant. Therefore, we cannot be fully confident in the estimates of contextual factors until we have accounted for the relative position of these places in space (for a similar discussion of place-based associations see Voss et al. 2006). Additionally, spatial consideration and the associated modeling techniques provide an avenue to assess processes connected to contagion or spatial diffusion, such as social acceptability. As noted in the previous section, accounting for social acceptability processes may be particularly relevant to contextual studies of health. Through my implementation of these methodological extensions and newly available software (i.e., HLM 7.0), I advance discussions of how place matters for maternal smoking and reinforce the necessity of incorporating spatially-informed approaches in contextual analyses.

METHODOLOGICAL DETAILS

Data

The individual-level data for this project come from restricted natality data with county identifiers that was supplied by the National Center for Health Statistics (2007). These data were linked to 2005-2009 American Community Survey (ACS) county estimates (US Census Bureau 2010). These data, including sample restrictions and the contextual unit of analysis, are ideal for this project because they have been used in recent research to demonstrate contextual associations with maternal smoking (Shoff and Yang 2013). Employing the same data and variables allow me to directly extend this work without concerns about differences in the underlying data.

The dependent variable is a dichotomous variable that is coded one for women who report smoking one or more cigarettes daily during pregnancy. Spatial modeling cannot yet accommodate binary outcomes, but my use of this individual-level dependent variable is not an issue since the level-2 outcome – the one relevant to the spatial modeling – is the random intercept, which is continuous. Given the comprehensive treatment of the variables in previous work (Shoff and Yang 2013) and my focus on the contribution of spatial processes to understanding contextual associations, I do not describe the other variables here and I limit discussion of them in the results section to differences between the baseline and spatial models. See Table 1 for a complete variable list. Finally, although closely aligned with previous research, I received separate institutional review board approval for this work.

[Table 1. Individual- and County-level Variables]

Due to software limitations, I am unable to conduct all components of the analysis using the full dataset (N level 1 = 3,318,295; N level 2 = 3,036). When attempting to run a spatial lag version of my HLM model the program (i.e., HLM 7.0) would end with an error message reporting that there was inadequate memory space to complete the model, even when using the

64-bit version and reduced samples (N level 1 = 394,878; N level 2 = 254). Therefore, in order to estimate the more complex models, I focus on counties in one state – Texas – and use a random sample of individuals stratified by county. Texas is a suitable choice because it has a racially and ethnically diverse population, as well as a large enough number of county units to estimate meaningful level-2 associations (N = 254). The stratified random sample of individuals was generated in Stata using the "sample" command combined with the "by" option (Stata Corp 2013). Alternative sample sizes were tried, but in order to maximize the number of individuals in each county while reducing the sample size sufficiently to allow for the model to converge, only up to 50 observations were selected for each county in Texas (N = 11,451).

HLM: Setting the Baseline

My first step is to establish the baseline contextual model using HLM 7.0 (Raudenbush et al. 2011). I specify a Bernouli distribution to accommodate my dichotomous dependent variable. All variables are uncentered and only the intercept is allowed to be random. This baseline model includes all individual- and county-level variables.

Spatial Regression and Diagnostics

A key contribution of this work is new information on how to estimate Moran's *I* statistics and spatial modeling diagnostics after estimating multilevel models in HLM 7.0 (Raudenbush et al. 2011). These spatially-informed tools are necessary for the appropriate use of spatial models and are not yet available in HLM 7.0 despite the addition of a spatial lag model option. I propose estimating these statistics in R with the "spdep" package (R Development Core Team 2012). In this analysis, I use a first-order queen contiguity spatial weights matrix to identify a county's neighbors, which is necessary for calculating both the Moran's *I* statistic and the spatial diagnostic tests. This specification is consistent with my expectation that the characteristics of nearby counties are most important for capturing spatial dependence related to understandings of smoking acceptability, and/or shared unobserved characteristics.

I estimate the spatial statistics in R using the level-2 residuals associated with a multilevel model. Although I develop this methodological contribution in reference to HLM 7.0, the underlying message is widely applicable – all models that employ spatially contiguous units of analysis need to acknowledge and assess the impact of spatial autocorrelation if we are to produce robust research and theory. The procedure for estimating the two types of statistics is similar, yet relies on slightly different files from HLM. For the Moran's *I* statistic, I import the residuals from the full model. Unfortunately, when estimating the spatial diagnostics, the final model residuals cannot be taken directly from HLM. This is because existing code for the spatial diagnostic tests requires a specific object type that requires that the level-2 portion of the model be conducted in R in order to estimate the spatial diagnostics. Therefore, for this second step, I import the level-2 residuals from an HLM model that includes only individual-level covariates so that I can estimate the level-2 portion of the model in R. These residuals have been purged of individual-level variation so they can be used to roughly approximate the level-2 results from HLM. However, I emphasize that this approach is not ideal for obtaining level-2 coefficient estimates because it is less efficient, and thus provides somewhat suspect estimates, particularly for the standard errors. That said, the residuals from this approach are statistically comparable to those from the full HLM model; therefore, conducting the level-2 model in R is an appropriate substitute for full-model residuals taken directly from HLM until new approaches are made available.

To assist with future assessment of the spatial structure in conjunction with multilevel modeling, I detail these steps below. However, the following steps assume some base knowledge of HLM 7.0 and R. Users new to these platforms should consult existing guides (see e.g., Hothorn and Everitt 2014; Raudenbush, Bryk, Cheong, Congdon, and du Toit 2011)

The first step is to create a spatial weights file. Starting from a shapefile matching the geographic units in your analysis, create a weights file using GeoDa or comparable software. The ID variable should preferably be in integer format. GeoDa will save the weights file in an ASCII format, but HLM 7.0 requires that the weights file be in the same format as the level-1 and level-2 data (e.g., SPSS format). In addition, the weights file needs to be set-up in a specific way in order for HLM 7.0 to read it correctly. To illustrate this reformatting process, I depict the adjustments needed to convert the original file (see Figure 1) into the HLM-accepted format (see Table 1; also see Raudenbush et al. 2011). First, I exported the ASCII file into Excel format using StatTransfer (Circle Systems, Inc. 2013). I chose Excel for its ease of data manipulation, but other software could have been used (e.g., Stata, SPSS). Second, I deleted the second row and renamed the columns to match those depicted in Table 2 (see Figure 1, Panel B). Third, I made a series of adjustments to the data so that each row represents a single level-2 observation and its neighbors. This requires moving the row of neighbors up to the row with the focal observation's ID number (see Figure 1, Panel C). Fourth, the number reported directly after the focal unit ID is the total number of neighbors for that unit, and this should be moved to the last column that is labeled "Count" (see Figure 1, Panel C). Finally, I exported the saved Excel file into SPSS format using StatTransfer.

[Figure 1. Starting Spatial Weights File Format]

[Table 2. HLM-friendly Spatial Weights File Format]

The second step is to generate the residual file used to estimate the Moran's *I* statistic. To start this process I set-up the full multilevel model that includes both individual- and county-level variables in HLM 7.0. Before running the analysis, I select the "Level-2 Residual File" option under "Basic Settings." This residual file will include a variety of variables automatically without prompting, including EBINTRCP – the focus of this analysis (for a full description of residual files in HLM see Taylor 2012). The EBINTRCP variable contains the residual estimates for each level-2 unit (i.e., U0J), and can therefore be used to assess residual spatial autocorrelation among the level-2 units as well as additional level-2 associations not already tested in the model. Before clicking "Ok," be sure to select your preferred file type, and change the suffix to correspond with your selection (e.g., the file name should end with ".dta" when "Stata" is selected). This file will be saved automatically to the same location as the HLM output after you run the model. Transfer this file into a format that can be easily read into R (e.g., Excel, .csv, or ASCII). Next, read the residual data file into R, install the "spdep" library to access the necessary code, and identify your spatial weights "list" object. Finally, estimate the Moran's *I* statistic for the EBINTRCP variable using the "moran.test" code. This will allow for an assessment of the extent to which the assumption of uncorrelated residuals is supported. A significant value, particularly values over 0.10, would suggest the need for further assessment of the residuals to determine which type of spatial model – error or lag – best represents the structure of the data.

The third step involves estimating the spatial diagnostics that are used to distinguish between spatial error and spatial lag data structures. This step follows closely with the previous step with the exception that the estimated model in HLM only includes the individual-level characteristics so that we can approximate the level-2 model in R. For ease, select all of the level-2 variables necessary for the model when creating the level-2 residual file so that everything you need for R is already combined into one file. After creating, transferring, and reading in this file, there are two steps. First, create an object from an OLS model where EBINTRCPT is the dependent variable and the independent variables are the same as those from the full model (e.g., model<- lm(EBINTRCPT~var1+var2…+var10, data=data_object)). Second, use the "lm.LMtests" code – also available through the "spdep" library in R – to run the diagnostic test on the "model" object. This will produce estimates relevant to distinguishing between an error and lag modeling approach.

Based on hypotheses related to social acceptability I derive two empirical expectations. First, consistent with a contagion or spatial diffusion process, I expect that a spatial lag model will be preferred. Second, I anticipate that contextual variables found to be significant when using standard multilevel models will no longer be significant after accounting for the spatial structure of the data, particularly those related to racial/ethnic composition. I demonstrate the utility of this spatially-informed HLM approach below using a contextual analysis of maternal smoking.

RESULTS

Contextual factors are clearly related to the individual-level odds of smoking, suggesting a role of place in shaping health outcomes. Focusing on the level-2 associations, the baseline model for Texas suggests that county SES, non-Hispanic black, and Hispanic population concentration are related to a woman's odds of smoking while pregnant (see Table 3). Counties with higher values on the SES scale have lower average odds of smoking during pregnancy. Similarly, a woman's odds – regardless of her own race – are much lower for every increase in the percent non-Hispanic black and percent Hispanic. This indicates that living in counties with relatively large black and Hispanic populations benefits reductions in maternal smoking.

These baseline findings are comparable to what Shoff and Yang (2013) report, yet in contrast, no other county-level association is significant. Fewer significant contextual variables may be a result of reduced statistical power, or it could be an indication that these associations differ across states and/or other social contexts. Although of less relevance to the current endeavor, it is reassuring that the individual-level associations reported here are also highly comparable with the results reported by Shoff and Yang (2013).

[Table 3. Baseline HLM Analysis, Texas]

Central to this analysis is the extent to which these baseline associations remain after accounting for any residual spatial processes. I assess the relevance of spatial autocorrelation for the baseline model using a Moran's *I* statistic. My results suggest that there is significant and substantively meaningful spatial autocorrelation in the level-2 residuals, even after accounting for key structural covariates and the distribution of individual characteristics across counties (I = 0.14, $p < .001$). This suggests that the results reported above, and in previous work (e.g., Shaw et al. 2010; Shaw and Pickett 2013; Shoff and Yang 2013), may be biased by correlated residuals.

But how should we characterize the underlying spatial structure? The diagnostic tests suggest that a spatial lag model is preferred over the spatial error specification (see Table 4). Both of the initial Lagrange multiplier estimates are significant, but between the robust estimates only the one for lag is significant at traditional levels. Given the proposed connection between contagion processes like social acceptability and a spatial lag manifestation, this result is consistent with arguments that suggest a role of social acceptability in explaining smoking (see especially Daly et al. 1993; Shaw and Pickett 2013).

[Table 4. Spatial Diagnostics on HLM Level-2 Residuals, Texas]

Further supporting a social acceptability argument, the results from the spatial multilevel model (conducted in HLM 7.0) indicate a positive spatial lag process – Rho is significant at the $p < 0.001$ level (see Table 5). Rho represents the association between the odds of smoking in one county and the average odds of maternal smoking in neighboring counties. Therefore, the positive valence suggests that the presence of the outcome in one county makes it more likely to be present in neighboring counties, net of individual and other county factors. Drawing from social acceptability arguments, the greater frequency of engagement in the act within the target population (i.e., smoking among pregnant women) in surrounding areas may contribute to, or at least signal, a sense of social acceptability and therefore a higher likelihood of an individual's own engagement.

[Table 5. Spatial HLM Results, Texas]

Related to this spatial diffusion process are the mediation of other contextual factors and the subsequent improvement of our contextual model of maternal smoking. Despite originally suggesting protective effects of black and Hispanic concentration, the results from the spatial version of the model show no such indication (see Table 5). After accounting for the spatial manifestation of social acceptability, racial/ethnic composition plays a limited independent role in explaining maternal smoking. These results bring new evidence to bear on the link between racial/ethnic composition and reduced smoking, and lend credence to the argument that social norms within black and Hispanic communities help to reduce the odds of smoking for all individuals in the area (see Shaw and Pickett 2013).

The persistence of contextual associations is also of note, as it suggests a truly robust relationship. The magnitude of all of the level-2 coefficients declined in the spatial model, but one remains strongly significant – the measure of SES (see Table 5). Supporting the high level of attention to this contextual characteristic and its accompanying manifestations, the persistence of the SES association after accounting for spatial processes suggests that it should remain a core component of contextual analyses of health.

DISCUSSION

Space and place are closely linked concepts that overlap and relate in many ways (see Gieryn 2000; Lobao 2004; Logan 2012). In fact, it is often difficult to distinguish between them, even within this analysis. Where does place end and space begin? As Logan (2012) argues, it is through reference to relative position that we are able to make this distinction. So, if one considers that neighboring places are relevant due to their geographic proximity, then the spatial layout is still centrally involved. Correspondingly, in this study, I demonstrate that the role of place cannot be accurately captured without considering its relative position among other places – that is, space.

My use of spatially-informed multilevel methods provides a robust foundation to inform our understanding of the contextual factors associated with maternal smoking during pregnancy. Ignoring space, particularly when analyzing spatially contiguous contextual units, could result in an overstatement, or misplacement, of the extent to which certain contextual characteristics matter. In this analysis, misplaced emphasis is demonstrated in the case of racial composition (discussed further below); and the non-significance of other factors (e.g., social capital and rurality) after accounting for spatial processes suggests an overstatement of their importance in previous work. Notably, I find a persistent association for the county SES indicator. This result provides strong support for a continued focus on community factors related to SES, particularly when examining health outcomes.

My incorporation of space also provides new insight into contextual relationships by examining the spatial manifestation of social acceptability processes. Like previous work, the results from the standard multilevel model would have indicated (vaguely) that black and Hispanic population concentration are related to lower odds of maternal smoking (Nkansah-Amankra 2010; Shaw et al. 2010; Shoff and Yang 2013). Although not definitive, the mediation of these associations in the spatial lag model suggests that they are actually the result of broader social acceptability processes (see especially Daly et al. 1993; Shaw and Pickett 2013). Future research can test the limits of this social acceptability explanation by addressing nuances that I am unable to capture, most notably the complex interplay between frequency and acceptability. It is likely that the level of acceptability affects levels of smoking and subsequent participation in smoking behaviors reinforces perceptions of acceptability. My analysis obscures the role of such feedback processes, but understanding this dynamic may prove central to addressing the health outcomes that are affected by social acceptability. An additional nuance relates to how we define “neighbors.” It may be necessary for research to consider social factors when identifying “neighbors” because proximity does not guarantee (or restrict) social influence. This extension is particularly relevant to spatial analyses that aim to capture social processes.

One more spatial consideration deserves note – the potential for spatial variation in relationships. Even when focusing on the baseline models, the results presented here for Texas indicate some differences from what was reported using the national sample (Shoff and Yang 2013). Given the use of an otherwise comparable model, this difference is suggestive of substantive differences within the United States in whether and how context is related to maternal smoking. Future research could empirically address this additional contextual layer using established approaches for assessing differences that unfold over space (see especially Baller et al. 2001; Curtis, Voss and Long 2012; Fotheringham, Brunsdon, and Charlton 2002). This consideration may be particularly important when developing policy recommendations as the relevance of a factor may vary across states or other identifiable contexts.

Finally, this study highlights gaps in existing software that may limit how space and place are combined in future research. Although beyond my own abilities, it is my hope that by bringing attention to these software issues and the constraints that they place on future knowledge development that those with the necessary skills will be called to action. I emphasize two main obstacles to the full utilization of recent advancements. First, the memory capacity of HLM 7.0 will need to be expanded in order to accommodate both the increasingly complex spatial HLM models, and the proliferation of large datasets. Second, ideally more spatial data analysis techniques would be available within HLM software. This would include the addition of Moran's *I*, spatial diagnostics, and spatial error regression estimation. The addition of these spatial tools will aid in the responsible use of spatial regression within the realm of HLM. However, the latter two additions may be particularly important given the different theoretical implications of the two spatial modeling strategies (i.e., spatial error versus lag).

The implications of this research reach across disciplines, subfields, and methods. Although my results can only speak directly to multilevel, quantitative approaches to studying place and context, the conceptual issues that I raise regarding the need to include space in order to understand place apply beyond this narrow focus. For instance, the influence of neighboring counties would still be reflected in a single county case study, and the inclusion of that information may affect our interpretation of how place and context matter. Therefore, the call to be critical about space and place applies more broadly (see Gieryn 2000; Lobao 2004; Logan 2012), but I demonstrate a particular need to reassess the persistence of contextual results – both null and significant – in light of spatial considerations. I assert that the incorporation of aspects of space will be a central feature of future theoretical advancements.

REFERENCES

Afifi, Rima, Joanna Khalil, Fouad Fouad, Fadi Hammal, Yara Jarallah, Hala Abu Farhat, Maha Ayad, and Rima Nakkash. 2013. "Social Norms and Attitudes Linked to Waterpipe use in the Eastern Mediterranean Region." *Social Science & Medicine* 98(December): 125-34.

Ajilore, Olugbenga, Aliaksandr Amialchuk, Wei Xiong, and Xinyue Ye. 2014. "Uncovering Peer Effects Mechanisms with Weight Outcomes using Spatial Econometrics." *Social Science Journal* 51(4): 645-51.

Albers, Allison B., Michael Siegel, Debbie M. Cheng, Lois Biener, and Nancy A. Rigotti. 2004. "Relation between Local Restaurant Smoking Regulations and Attitudes towards the Prevelance and Social Acceptability of Smoking: A Study of Youths and Adults who eat Out Predominantly at Restaurants in their Town." *Tobacco Control* 13: 347-55.

Alesci, Nina L., Jean L. Forster, and Therese Blaine. 2003. "Smoking Visibility, Perceived Acceptability, and Frequency in Various Locations among Youth and Adults." *Preventive Medicine* 36(3): 272-81.

Baller, Robert D., Luc Anselin, Steven F. Messner, Glenn Deane, Darnell F. Hawkins. 2001. "Structural Covariates of U.S. County Homicide Rates: Incorporating Spatial Effects." *Criminology* 39: 561-590.

Boardman, Jason D. 2004. "Stress and Physical Health: The Role of Neighborhoods as Mediating and Moderating Mechanisms." *Social Science & Medicine* 58(12): 2473-83.

Botvin, Gilbert J., Linda Dusenbury, Eli Baker, Susan James-Ortiz, Elizabeth M. Botvin, and Jon Kerner. 1992. "Smoking Prevention among Urban Minority Youth: Assessing Effects on Outcome and Mediating Variables." *Health Psychology* 11(5): 290-9.

Brooks-Gunn, Jeanne, Greg J. Duncan, and J. Lawrence Aber, Eds. 1997. *Neighborhood Poverty, Volume I: Context and Consequences for Children*. New York: Russell Sage Foundation.

Bryk, Anthony S. and Stephen W. Raudenbush. 1992. *Hierarchical Models: Applications and Data Analysis Methods*. Sage.

Chaix, B., Merio, J., & Chauvin, P. 2004. Investigating place effects on health: A spatial approach v the multilevel approach. *Journal of Epidemiology & Community Health* 58(Aug), A14.

Circle Systems, Inc. 2013. *StatTransfer 12*.

Clarke, Philippa, Jeffrey Morenoff, Michelle Debbink, Ezra Golberstein, Michael R. Elliott, and Paula M. Lantz. 2014. "Cumulative Exposure to Neighborhood Context: Consequences for Health Transitions Over the Adult Life Course." *Research on Aging* 36(1): 115-42.

Cliff, A. D., Ord, J. K. 1973. *Spatial Autocorrelation*. Pion Limited, London, England.

Cliff, A. D., Ord, J. K. 1981. *Spatial Processes: Models & Applications*. Pion Limited, London, England.

Crowder, Kyle and Scott J. South. 2008. "Spatial Dynamics of White Flight: The Effects of Local and Extralocal Racial Conditions on Neighborhood Out-Migration." *American Sociological Review* 73(5): 792-812.

Crowder, Kyle and Scott J. South. 2011. "Spatial and Temporal Dimensions of Neighborhood Effects on High School Graduation." *Social Science Research* 40(1): 87-106.

Curtis, Katherine J., Paul R. Voss, and David D. Long. 2012. "Spatial Variation in Poverty-Generating Processes: Child Poverty in the United States." *Social Science Research* 41(1): 146-159.

Daly, Kathleen A., Elizabeth M. Lund, Kathleen C. Harty, and Susan A. Ersted. 1993. "Factors Associated with Late Smoking Initiation in Minnesota Women." *American Journal of Public Health* 83(9): 1333-5.

Economic Research Service. 2003. *Rural-urban Continuum Codes*. United States: Department of Agriculture.

Fotheringham, A. Stewart, Chris Brunsdon, and Martin Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. London: Wiley.

Gieryn, Thomas F. 2000. "A Space for Place in Sociology." *Annual Review of Sociology* 26: 463-96.

Hothorn, Torsten and Brian S. Everitt. 2014. *A Handbook of Statistical Analyses using R*. Boca Raton, FL: CRC Press.

Hruschka, Daniel J., Alexandra A. Brewis, Amber Wutich, and Benjamin Morin. 2011. "Shared Norms and their Explanation for the Social Clustering of Obesity." *American Journal of Public Health* 101(S1): S295-300.

Kimbro, Rachel T. 2009. "Acculturation in Context: Gender, Age at Migration, Neighborhood Ethnicity, and Health Behaviors." *Social Science Quarterly* 90: 1145-66.

Kimbro, Rachel T. and Justin T. Denney. 2013. "Neighborhood Context and Racial/Ethnic Differences in Young Children's Obesity: Structural Barriers to Interventions." *Social Science & Medicine* 95: 97-105.

Landman, Anne, Daniel K. Cortese, and Stanton Glantz. 2008. "Tobacco Industry Sociological Programs to Influence Public Beliefs about Smoking." *Social Science & Medicine* 66(2008): 970-81.

Lobao, Linda. 2004. "Continuity and Change in Place Stratification: Spatial Inequality and Middle-Range Territorial Units." *Rural Sociology* 69(10: 1-30.

Mu, Lan, Fahui Wang, Vivien W. Chen, and Xiao-Cheng Wu. 2015. "A Place-Oriented, Mixed-Level Regionalization Method for Constructing Geographic Areas in Health Data Dissemination and Analysis." *Annals of the Association of American Geographers* 105(1): 48-66.

National Center for Health Statistics. 2007. Machine readable data file and documentation, CD-ROM. In *National vital statistics system detail natality files*. Hyattsville, MD: US Department of Health and Human Services.

Nkansah-Amankra, Stephen. 2010. "Neighborhood Contextual Factors, Maternal Smoking, and Birth Outcomes: Multilevel Analysis of the South Carolina PRAMS Survey, 2000-2003." *Journal of Women's Health* 19(8): 1543-52.

Patterson, J. M., Eberly, L. E., Ding, Y., & Hargreaves, M. 2004. Associations of smoking prevalence with individual and area level social cohesion. *Journal of Epidemiology and Community Health* 58(8), 692-7.

Perchoux, C., Kestens, Y., Thomas, F., Hulst, A., Thierry, B., & Chaix, B. 2014. Assessing patterns of spatial behavior in health studies: their socio-demographic determinants and associations with transportation modes (the RECORD Cohort Study). *Social Science & Medicine* 119, 64-73.

R Development Core Team. 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria.

Raudenbush, Stephen W., Anthony S. Bryk, and Richard Congdon. 2011. *HLM 7 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.

Raudenbush, Stephen W., Anthony S. Bryk, Yuk Fai Cheong, Richard T. Congdon, Jr., and Mathilda du Toit. 2011. "HLM 7: Hierarchical Linear and Nonlinear Modeling." Lincolnwood, IL: Scientific Software International, Inc.

Rupasingha, A., S. J. Goetz, and D. Freshwater. 2006. "The Production of Social Capital in US Counties." *Journal of Socio-economics* 35(1): 83-101.

Shaw, Richard J. and Kate E. Pickett. 2013. "The Health Benefits of Hispanic Communities for Non-Hispanic Mothers and Infants: Another Hispanic Paradox." *American Journal of Public Health* 103(6): 1052-7.

Shaw, Richard J., Kate E. Pickett, and Richard G. Wilkinson. 2010. "Ethnic Density Effects on Birth Outcomes and Maternal Smoking during Pregnancy in the US Linked Birth and Infant Death Data Set." *American Journal of Public Health* 100(4): 707-13.

Shoff, Carla and T.C. Yang. 2013. "Understanding Maternal Smoking During Pregnancy: Does Residential Context Matter?" *Social Science & Medicine* 78(2013): 50-60.

Stata Corp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP.

Taylor, Ralph B. 2012. "Residual Files in HLM." Retrieved January 14, 2015 from http://www.rbtaylor.net/HLML09_20120119.pdf.

Teachman, Jay and Kyle Crowder. 2002. "Multilevel Models in Family Research: Some Conceptual and Methodological Issues." *Journal of Marriage & Family* 64(2): 280-94.

Thomson, Carey Conley, Michael Siegel, Jonathan Winickoff, Lois Biener, and Nancy A. Rigotti. 2005. "Household Smoking Bans and Adolescents' Perceived Prevalence of Smoking and Social Acceptability of Smoking." *Preventive Medicine* 41(2): 349-56.

US Census Bureau. 2010. American Community Survey, United States. Retrieved March, 12 2014 from http://factfinder.census.gov.

Verbitsky-Savitz, Natalya and Stephen W. Raudenbush 2009. "Exploiting Spatial Dependence to Improve Measurement of Neighborhood Social Processes." *Sociological Methodology* 39(1): 151-83.

Voss, Paul R., David D. Long, Roger B. Hammer, and Samantha Friedman. 2006. "County Child Poverty Rates in the US: A Spatial Regression Approach." *Population Research and Policy Review* 25(4): 369-91.

Xu, Hongwei, John R. Logan, and Susan E. Short. 2014. "Integrating Space with Place in Health Research: A Multilevel Spatial Investigation using Child Mortality in 1880 Newark, New Jersey." *Demography* 51(3): 811-34.

Yang, Tse-Chuan and Stephen A. Matthews. 2010. "The Role of Social and Built Environments in Predicting Self-rated Stress: A Multilevel Analysis in Philadelphia." *Health and Place* 16: 803-10.

Yang, Tse-Chuan, Stephen A. Matthews, and Carla Shoff. 2011. "Individual Health Care System Distrust and Neighborhood Social Environment: How are they Jointly Associated with Self-rated Health? *Journal of Urban Health* 88(5): 945-58.

Table 1. Individual- and County-level Variables

Individual-level	
Smokes	A dichotomous variable coded one for women who smoked during pregnancy
Age (non-linear)	A continuous measure of age combined with a squared term
Race	A series of binary variables coded for the major racial categories in the United States (i.e., white, black/African American, American Indian/Alaskan native, and Asian), where white is used as the reference category
Hispanic	A binary variable coded one for women of any race who is Hispanic
Married	A dichotomous variable coded one for women who were married at the time of birth
Education	A series of binary variables coded for a woman's highest level of educational attainment at the time of the birth – less than high school (ref), high school/GED, some college/associate's degree, and bachelor's degree or higher
Weight Gain (non-linear)	A continuous variable of how much weight a woman gained during pregnancy combined with a squared term
Prenatal care	A series of binary variables that categorizes the adequacy of a woman's prenatal care accounting to the timing and number of visits – inadequate (ref), intermediate, adequate, and adequate plus care
Parity	A dichotomous variable coded one if the 2007 birth were the woman's first birth
County-level	
Rural	A binary variable coded one if a county had codes 8 or 9 on the US Department of Agriculture Economic Research Service Rural-Urban Continuum (Economic Research Service 2003)
Socioeconomic Status	A composite measure derived using principal component analysis and the following variables: per capita income; percent with a bachelor's degree; percent employed in professional, administrative, and managerial positions; percent of families with an income above 75,000 dollars; percent of families in poverty; and percent of families with a female head and children under 18
Social Capital	An index measure originally developed by Rupasingha et al. (2006)
Social Capital Interaction	The social capital index multiplied by the rural binary variable
Racial/ethnic Composition	Three variables are included to describe the composition of the total population: the percent non-Hispanic white, the percent non-Hispanic black/African American, and the percent Hispanic

Table 2. HLM-friendly Spatial Weights File Format

id	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	count
48295	48211	48393	48357								3
48421	48341	48205	48111	48195	48233						5
48111	48205	48421	48341								3
48195	48341	48233	48357	48393	48421						5

Table 3. Baseline HLM Analysis, Texas

	Odds Ratio
Intercept, γ_{00}	0.92
Percent Non-Hispanic White	0.02
Percent Non-Hispanic Black[a]	0.01*
Percent Hispanic[a]	0.01*
SES	0.73***
Social Capital Index	1.03
Rural	1.08
Social Capital-Rural Interaction	0.93
Random Effect, υ_0	0.10***
Age	1.40***
Age^2	0.99***
White	(ref)
Black	0.37***
American Indian/Alaskan Native	0.52
Asian	0.28**
Non-Hispanic	(ref)
Hispanic	0.13***
Married	0.49***
Less than High School	(ref)
High School/GED	0.69***
Some College/Associate’s	0.39***
Bachelor’s or Higher	0.04***
Weight Gain	0.99**
Weight Gain^2	1.00***
Inadequate Prenatal Care	(ref)
Intermediate	0.68**
Adequate	0.64***
Adequate Plus	0.68***
First Birth	0.76***

Note: Coefficients are reported as odds ratios.
Significance is based on robust standard errors.
[a] Odds ratios were less than 0.01.

Table 4. Spatial Diagnostics on HLM Level-2 Residuals, Texas

Spatial Dependence Structure	Lagrange Multiplier
Error	9.43**
Lag	15.46***
Error (robust)	3.33†
Lag (robust)	9.36**

Table 5. Spatial HLM Results, Texas

	Odds Ratio
Intercept, γ_{00}	0.06
Percent Non-Hispanic White	0.34
Percent Non-Hispanic Black	0.12
Percent Hispanic	0.04
SES	0.77***
Social Capital Index	0.91
Rural	1.10
Social Capital-Rural Interaction	0.95
Spatial Lag (Rho)	0.90***

Note: Individual coefficients are unchanged from baseline.

Figure 1. Starting Spatial Weights File Format

Panel A

Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
0	254	file	id var					
48295	3							
48211	48393	48357						
48421	5							
48341	48205	48111	48195	48233				
48111	3							
48205	48421	48341						
48195	5							
48341	48233	48357	48393	48421				

Panel B

id	n1	n2	n3	n4	n5	n6	n7	n8	***n9***	***n10***	***Count***
0	254	file	id var								
48295	3										
48211	48393	48357									
48421	5										
48341	48205	48111	48195	48233							
48111	3										
48205	48421	48341									
48195	5										
48341	48233	48357	48393	48421							

Panel C

id	n1	n2	n3	n4	n5	n6	n7	n8	***n9***	***n10***	***Count***
48295	$\boxed{3}$										
48211	48393	48357									
48421	$\boxed{5}$										
48341	48205	48111	48195	48233							
48111	$\boxed{3}$										
48205	48421	48341									
48195	$\boxed{5}$										
48341	48233	48357	48393	48421							