Published in final edited form as:

Soc Sci Med. 2015 September ; 140: 18–26. doi:10.1016/j.socscimed.2015.06.025.

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, 6100 Main Street, MS-208, Houston, TX 77005, Phone: 713-348-2986

Heather A. O'Connell: hoconnell@rice.edu

Abstract

I contribute to understandings of how context is related to 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. For this analysis I use restricted-use natality data from the Vital Statistics, and county-level data from the 2005–9 ACS. Critically, the results suggest that spatial considerations are still relevant in a multilevel framework. In addition, I argue that spatial processes help explain the relationships linking racial/ethnic minority concentration to lower overall odds of smoking.

Keywords

space; place; multilevel modeling; spatial modeling; health; context; racial/ethnic composition; United States

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 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 resulting conceptual and methodological benefits. In addition, I extend previous research that combines multilevel and spatial modeling by demonstrating how to obtain Moran's *I* estimates and spatial model diagnostics – necessary,

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

yet unavailable (in the multilevel context), guiding tools. Given the co-dependence of place and space, these and related methodological advancements are central to our pursuit of a more sophisticated and realistic understanding of how both are involved in the production of 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 Raudenbush 1992). Using a similar logic, spatial models address theories and statistical issues related to how spatial position affects social processes and statistical estimates (Cliff and Ord 1973, 1981). When we do not account for the clustering of individual observations and spatial positioning, we have a higher risk of coming to the wrong conclusion than otherwise expected (Bryk and Raudenbush 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 contextual relationships – however defined – cannot be properly assessed without considering concepts related to space 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. This is an ideal focus because research suggests not only that there are strong connections between county context and the odds of smoking during pregnancy (Shoff and Yang 2013); it also suggests the role of social acceptability in smoking behaviors, a social process that 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). Critically, my application of this methodological extension suggests that previous work using multilevel modeling alone may have overestimated contextual associations. In addition, the results extend the scope of previous work on maternal smoking by assessing the social acceptability hypothesis, and by more accurately identifying contextual factors relevant for explaining smoking.

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 and Denney 2013; Yang and Matthews 2010; Morenoff 2003; Taylor, Repetti, and Seeman 1997). The primary contextual factors examined in many multilevel studies 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. Yet, it is not the only local characteristic that requires attention.

Of most relevance to the present research is the significance and interpretation of racial composition variables. Previous research indicates 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. I discuss explanations for this association below.

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, related to the results for racial/ethnic composition, I highlight a difficult to capture process that is of potential wide-reaching significance – social acceptability.

THE ELUSIVE LINK BETWEEN SOCIAL ACCEPTABILITY AND SMOKING

The topic of social acceptability has entered into a range of health research. Centrally, 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; 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). Beyond smoking, 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); arguably, ideas about what is acceptable may be related to even more health behaviors, like utilizing doctors. Despite suggestion 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 her 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 other 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 spatial processes, including those potentially tied to social acceptability. Detailing the specifics of acceptability processes will be a central challenge for future research on health. But to stay focused on the task at hand – extending multilevel frameworks through spatial considerations – 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. 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. After identifying spatial autocorrelation among model residuals using the Moran's *I* statistic, researchers use diagnostic tools to determine the most appropriate spatial model – either the spatial lag or spatial error. Although the Moran's *I* statistic is most closely linked to an error interpretation of spatial processes given its focus on the model residuals, the information from this statistic is critical to identifying when other spatial processes, such as those related to a spatial lag, may also be involved because both would result in spatially correlated residuals.

Despite both suggesting a statistical link among neighbors, these two approaches to spatial modeling – lag and error – 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 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.

Spatial considerations have been incorporated directly into multilevel research in a variety of ways, but they 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 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). For instance, 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.

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. And I emphasize that the correlated residuals may be due either to lag or error processes. 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 considerations 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. I note that, although closely aligned with previous research, I received separate institutional review board approval for this work.

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.

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 multilevel model, HLM 7.0 would give an error message reporting that there was inadequate memory space to complete the model, even when using the 64-bit version. 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 ($N = 254$). Fortunately, despite some differences in educational attainment, prenatal care, and the expected racial/ethnic differences associated with a focus on Texas, the descriptive statistics suggest that my individual sample is otherwise comparable to what was analyzed by Shoff and Yang (2013) – e.g., nearly 12 percent of the women in my analysis were coded as smokers whereas 11 percent was reported for the national data (full descriptive information available upon request). 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 (e.g., $N = 394,878$), 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 Bernoulli 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 spatial weights specification is consistent with my expectation that the characteristics of nearby counties are most important for capturing spatial dependence related to smoking acceptability (i.e., spatial lag/spatial diffusion), and/or shared unobserved characteristics (i.e., spatial error).

I estimate the spatial statistics (i.e., Moran's I and spatial model diagnostics) in R using the level-2 residuals associated with the 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 of the data in conjunction with multilevel modeling, I detail the necessary 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. 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, Panel A) into the HLM-accepted format (see Raudenbush et al. 2011). First, I exported the ASCII file created using GeoDa software into an Excel format using StatTransfer (Circle Systems, Inc. 2013). Note that in all cases other software could have been used (e.g., R instead of GeoDa, and Stata instead of Excel). Second, I deleted the second row and renamed the columns to match the HLM-accepted format (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 (i.e., move the row with the codes for the neighbors up a row). 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, 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). Therefore, I exported the saved Excel file into SPSS format using StatTransfer.

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, 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 because correlated residuals could be due to the omission of spatially clustered variables (i.e., spatial error), *or* more specifically due to a relationship between neighboring units’ outcomes (i.e., spatial diffusion/lag).

The third step involves estimating the spatial diagnostics that are used to determine which type of spatial model – error or lag – best represents the structure of the data. 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 – 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. However, I emphasize that while consistent with the social acceptability hypothesis, a spatial lag model is an indirect, spatial assessment of this social process. 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 the variables 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 2). 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 is beneficial for reducing maternal smoking.

These three significant, county-level findings are comparable to what Shoff and Yang (2013) report, as are the individual-level associations. However, in contrast, my results suggest no other county-level association is significant. Fewer significant contextual variables may be a result of reduced statistical power. Alternatively, and of greater substantive interest, the differences could be an indication that associations vary across states and/or other social contexts. I pick up on this point again in the discussion.

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. Furthermore, this result provides the impetus for thinking about the spatial processes driving this residual spatial autocorrelation – is it strictly error based, or could processes related to spatial diffusion be involved?

The diagnostic tests suggest that a spatial lag model is preferred over the spatial error specification, which implicates spatial diffusion processes (see Table 3). Both of the initial Lagrange multiplier estimates are significant, but between the robust estimates – where a robust estimate for lag refers to one that is robust to existing error processes, and vice versa – only the one for lag is significant at traditional levels. These results suggest that, although not the only spatial process involved, processes related to a spatial lag are more dominant. 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).

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 4). Rho represents the association between maternal smoking in one county and the average in neighboring counties. Therefore, the positive valence suggests that the presence of the outcome in one county makes it more likely to occur in neighboring counties, net of individual and other county factors. Drawing from social acceptability arguments, the greater frequency of engagement in the act in surrounding areas may contribute to, or at least signal, a sense of social acceptability that then contributes to a higher likelihood of an individual's own engagement in the behavior.

Finally, I turn to an assessment of how the inclusion of Rho affects the interpretation of the contextual covariates. First, the SES contextual association is virtually unchanged after accounting for spatial processes, which suggests a truly robust relationship with maternal smoking. Second, and in contrast, 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 4). However, I note that this mediating result may be most applicable when considering the black concentration relationship. When excluding the non-Hispanic

white concentration and social capital-rural interaction variables – two variables that are highly correlated with other variables in the model – the spatial lag coefficient remains the same, but there are two notable changes. First, as was the case in the baseline model, Hispanic concentration remains significant ($p < .001$; not shown). Second, the social capital index is marginally related to a lower average odds of maternal smoking, but only in the spatial multilevel model ($p < .06$; not shown). The implications of and questions raised by these results are discussed below.

DISCUSSION

Place and space are closely linked concepts that overlap and relate in many ways (see Gieryn 2000; Lobao 2004; Logan 2012). Although often difficult to distinguish, both are necessary for explaining how social processes unfold. The incorporation of aspects of space and place has proliferated within the social sciences, but we are less comfortable with addressing issues related to where they intersect. For example, how does space affect our understanding of associations attributed to place? In this study, I demonstrate that the role of place cannot be accurately, or even fully, captured without considering its relative position among other places – that is, space.

My use of spatially-informed multilevel methods provides a stronger foundation for understanding the contextual factors associated with maternal smoking during pregnancy than was previously available. 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., rurality), suggests an overstatement or overgeneralization of their importance in previous work. Notably, in the face of all of this non-significance, I find a persistent association for the county SES indicator. This result provides support for the continued focus on community factors related to SES, particularly when examining health outcomes.

My incorporation of space also provides 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. This finding corroborates previously untested arguments (see especially Daly et al. 1993; Shaw and Pickett 2013), and contributes to intervention development by suggesting a focus on perceptions of what is acceptable.

Social acceptability is a potentially powerful explanation for smoking patterns, but it may be less relevant for explaining the Hispanic concentration association, at least when focusing on diffusion at the county scale. This could suggest that there are additional smoking-deterrent dynamics related to Hispanic concentration. However, there are also statistical issues to consider before dismissing a social acceptability explanation for the Hispanic concentration

association. The change in significance for Hispanic concentration in the spatial model after excluding the concentration of non-Hispanic whites suggests that multicollinearity may be an issue when examining racial/ethnic context using percentage variables. This statistical issue, and a related conceptual question regarding what is most important about the racial/ethnic context, could be addressed by using categorical variables that define places based on which group represents the majority, or some other alternative characterization. Identifying the process(es) underlying the strong association with racial/ethnic composition could be pivotal in shaping future intervention efforts.

Future research can test the limits of the social acceptability explanation by addressing nuances that I am unable to capture – most notably the complex interplay between frequency and acceptability, and manifestations at smaller scales – and alternative explanations for the spatial lag results. It is likely that the level of acceptability affects levels of smoking, and subsequent participation in smoking behaviors then 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 alongside proximity because proximity does not guarantee (or restrict) social influence. Related, research could consider more localized geographies that may better capture the social spaces relevant for the diffusion of ideas. Although the county unit was ideal from a consistency standpoint, subsequent efforts focused explicitly on social acceptability should consider additional scales.

Finally, despite being suggestive of a social contagion, social acceptability is not the only explanation for my spatial lag results. Additional analyses that employ a more direct measure of social acceptability are necessary to fully assess this explanation. However, baring a direct measure, research should assess alternative explanations including hypotheses related to the spatial clustering of SES factors, as well as hypotheses related to larger scale processes. Regarding the latter, it is possible that maternal smoking is clustered in neighboring counties as a result of regional norms. Assessing this possibility is important because regional norms could generate spatial patterns similar to those associated with diffusion processes, but would not implicate active spatial diffusion as an explanation for the pattern.

Turning back to my discussion of the role of space, 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 relationship 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 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 ability to estimate Moran's *I* statistics, spatial diagnostics, and spatial error regression models. The addition of all three 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).

Building from recent calls to be critical about space and place (see Gieryn 2000; Lobao 2004; Logan 2012), I demonstrate a particular need to reassess the persistence of contextual, multilevel results – both null and significant – in light of spatial considerations. 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. Correspondingly, I expect that this push to incorporate aspects of space will be a central feature of future theoretical advancements across disciplines, subtopics, and even methods.

Acknowledgments

This research was supported by center Grant #R24 HD047873 and training Grant #T32 HD07014 awarded to the Center for Demography and Ecology at the University of Wisconsin at Madison. Additional support was received through the Kinder Institute for Urban Research at Rice University. I would also like to specifically thank Carla Shoff, Justin Denney, and Katherine Curtis for their support while producing this work.

References

- Afifi R, Khalil J, Fouad F, Hammal F, Jarallah Y, Abu Farhat H, Ayad M, Nakkash R. Social norms and attitudes linked to waterpipe use in the eastern Mediterranean region. *Social Science & Medicine*. 2013; 98(December):125–134. [PubMed: 24331890]
- Ajilore O, Amialchuk A, Xiong W, Ye X. Uncovering peer effects mechanisms with weight outcomes using spatial econometrics. *Social Science Journal*. 2014; 51(4):645–651.
- Albers AB, Siegel M, Cheng DM, Biener L, Rigotti NA. Relation between local restaurant smoking regulations and attitudes towards the prevalence and social acceptability of smoking: a study of youths and adults who eat out predominantly at restaurants in their town. *Tobacco Control*. 2004; 13:347–355. [PubMed: 15564617]
- Alesci NL, Forster JL, Blaine T. Smoking visibility, perceived acceptability, and frequency in various locations among youth and adults. *Preventive Medicine*. 2003; 36(3):272–281. [PubMed: 12634018]
- Baller RD, Anselin L, Messner SF, Deane G, Hawkins DF. Structural covariates of U.S. county homicide rates: incorporating spatial effects. *Criminology*. 2001; 39:561–590.
- Boardman JD. Stress and physical health: the role of neighborhoods as mediating and moderating mechanisms. *Social Science & Medicine*. 2004; 58(12):2473–2483. [PubMed: 15081198]

- Botvin, Gilbert J.; Dusenbury, L.; Baker, E.; James-Ortiz, S.; Botvin, EM.; Kerner, J. Smoking prevention among urban minority youth: assessing effects on outcome and mediating variables. *Health Psychology*. 1992; 11(5):290–299. [PubMed: 1425546]
- Brooks-Gunn, J.; Duncan, GJ.; Aber, JL., editors. *Neighborhood poverty, volume I: context and consequences for children*. New York: Russell Sage Foundation; 1997.
- Bryk, AS.; Raudenbush, SW. *Hierarchical models: applications and data analysis methods*. Sage; 1992.
- Circle Systems, Inc. *StatTransfer*. 2013; 12
- Clarke P, Morenoff J, Debbink M, Golberstein E, Elliott MR, Lantz PM. Cumulative exposure to neighborhood context: consequences for health transitions over the adult life course. *Research on Aging*. 2014; 36(1):115–142. [PubMed: 24465068]
- Cliff, AD.; Ord, JK. *Spatial autocorrelation*. Pion Limited; London, England: 1973.
- Cliff, AD.; Ord, JK. *Spatial processes: models & applications*. Pion Limited; London, England: 1981.
- Crowder K, South SJ. Spatial dynamics of white flight: the effects of local and extralocal racial conditions on neighborhood out-migration. *American Sociological Review*. 2008; 73(5):792–812. [PubMed: 20221414]
- Crowder K, South SJ. Spatial and temporal dimensions of neighborhood effects on high school graduation. *Social Science Research*. 2011; 40(1):87–106. [PubMed: 21180398]
- Curtis KJ, Voss PR, Long DD. Spatial variation in poverty-generating processes: child poverty in the United States. *Social Science Research*. 2012; 41(1):146–159. [PubMed: 23017703]
- Daly KA, Lund EM, Harty KC, Ersted SA. Factors associated with late smoking initiation in Minnesota women. *American Journal of Public Health*. 1993; 83(9):1333–1335. [PubMed: 8363013]
- Economic Research Service. *Rural-urban Continuum Codes*. United States: Department of Agriculture; 2003.
- Fotheringham, AS.; Brunsdon, C.; Charlton, M. *Geographically weighted regression: the analysis of spatially varying relationships*. London: Wiley; 2002.
- Gieryn TF. A Space for Place in Sociology. *Annual Review of Sociology*. 2000; 26:463–496.
- Hothorn, T.; Everitt, BS. *A handbook of statistical analyses using R*. Boca Raton, FL: CRC Press; 2014.
- Hruschka DJ, Brewis AA, Wutich A, Morin B. Shared Norms and their Explanation for the Social Clustering of Obesity. *American Journal of Public Health*. 2011; 101(S1):S295–S300. [PubMed: 21555656]
- Kimbro RT, Denney JT. Neighborhood context and racial/ethnic differences in young children's obesity: structural barriers to interventions. *Social Science & Medicine*. 2013; 95:97–105. [PubMed: 23089614]
- Lobao L. Continuity and change in place stratification: spatial inequality and middle- range territorial units. *Rural Sociology*. 2004; 69(10):1–30.
- Mu L, Wang F, Chen VW, Wu X. A place-oriented, mixed-level regionalization method for constructing geographic areas in health data dissemination and analysis. *Annals of the Association of American Geographers*. 2015; 105(1):48–66. [PubMed: 26251551]
- National Center for Health Statistics. *National vital statistics system detail natality files*. Hyattsville, MD: U.S. Department of Health and Human Services; 2007. Machine readable data file and documentation, CD-ROM.
- Nkansah-Amankra S. Neighborhood contextual factors, maternal smoking, and birth outcomes: multilevel analysis of the South Carolina prams survey, 2000–2003. *Journal of Women's Health*. 2010; 19(8):1543–1552.
- Perchoux C, Kestens Y, Thomas F, Hulst A, Thierry B, Chaix B. Assessing patterns of spatial behavior in health studies: their socio-demographic determinants and associations with transportation modes (the RECORD Cohort Study). *Social Science & Medicine*. 2014; 119(1):64–73. [PubMed: 25150652]
- R Development Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing; Vienna, Austria: 2012.

- Raudenbush, SW.; Bryk, AS.; Congdon, R. HLM 7 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc; 2011.
- Raudenbush, SW.; Bryk, AS.; Cheong, YF.; Congdon, RT.; du Toit, M. HLM 7: Hierarchical Linear and Nonlinear Modeling. Lincolnwood, IL: Scientific Software International, Inc; 2011.
- Shaw RJ, Pickett KE. The health benefits of Hispanic communities for non- Hispanic mothers and infants: another Hispanic paradox. *American Journal of Public Health*. 2013; 103(6):1052–1057. [PubMed: 23597369]
- Shaw RJ, Pickett KE, Wilkinson RG. 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*. 2010; 100(4):707–13. [PubMed: 20167891]
- Shoff C, Yang TC. Understanding Maternal Smoking During Pregnancy: Does Residential Context Matter? *Social Science & Medicine*. 2013; 78(2013):50–60. [PubMed: 23246395]
- Stata Corp. Stata statistical software: release 13. College Station, TX: StataCorp LP; 2013.
- Taylor, RB. Residual Files in HLM. 2012. Retrieved January 14, 2015 from http://www.rbtaylor.net/HLM09_20120119.pdf
- Teachman J, Kyle Crowder. Multilevel models in family research: some conceptual and methodological issues. *Journal of Marriage & Family*. 2002; 64(2):280–294.
- Thomson CC, Siegel M, Winickoff J, Biener L, Rigotti NA. Household smoking bans and adolescents' perceived prevalence of smoking and social acceptability of smoking. *Preventive Medicine*. 2005; 41(2):349–356. [PubMed: 15917032]
- US Census Bureau. American Community Survey, United States. 2010. Retrieved March, 12 2014 from <http://factfinder.census.gov>
- Verbitsky-Savitz N, Raudenbush SW. Exploiting spatial dependence to improve measurement of neighborhood social processes. *Sociological Methodology*. 2009; 39(1):151–183.
- Voss PR, Long DD, Hammer RB, Friedman S. County child poverty rates in the US: a spatial regression approach. *Population Research and Policy Review*. 2006; 25(4):369–391.
- Xu H, Logan JR, Short SE. Integrating space with place in health research: a multilevel spatial investigation using child mortality in 1880 Newark, New Jersey. *Demography*. 2014; 51(3):811–834. [PubMed: 24763980]
- Yang TC, Matthews SA. The role of social and built environments in predicting self-rated stress: a multilevel analysis in Philadelphia. *Health and Place*. 2010; 16:803–810. [PubMed: 20434389]

HIGHLIGHTS

- I extend conceptual links between place and space to advance multilevel modeling.
- Additional spatial statistics improve our use of spatial modeling in this context.
- Consistent with social acceptability, spatial diffusion typifies maternal smoking.
- The role of racial/ethnic composition in smoking is explained by spatial diffusion.
- Incorporating space is central to enhancing theory on contextual linkages.

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	3										
48211	48393	48357									
48421	5										
48341	48205	48111	48195	48233							
48111	3										
48205	48421	48341									
48195	5										
48341	48233	48357	48393	48421							

Figure 1.
Initial Spatial Weights File Format and Conversion Steps

Table 1**Variable Descriptions**

<u>Individual-level</u>	
Smokes	Dichotomous variable coded one for women who smoked one or more cigarettes daily during pregnancy
Age (non-linear)	Continuous measure of age and its squared value
Race	Series of binary variables for white (ref), black/African American, American Indian/Alaskan native, and Asian
Hispanic	Women of any race who are Hispanic are coded as one
Married	Women who were married at the time of birth are coded as one
Education	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)	Continuous variable of how much weight a woman gained during pregnancy and its squared value
Prenatal care	Series of binary variables for a woman's prenatal care – inadequate (ref), intermediate, adequate, and adequate plus care
Parity	Dichotomous variable coded one if the 2007 birth were the woman's first birth
<u>County-level</u>	
Rural	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	Composite measure of local SES derived using principal component analysis
Social Capital	Index measuring the social connectedness of residents within a county
Social Capital-Rural Interaction	Social capital index multiplied by the binary rural 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

Note: See Shoff and Yang (2013) for complete details on variable construction.

Table 2

Baseline HLM Analysis, Texas

	<u>Odds Ratio</u>
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, ν_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: Significance is based on robust standard errors.

^a Odds ratios were less than 0.01.

*
p < 0.05,

**
p < 0.01,

p < 0.001

Table 3

Spatial Diagnostics on HLM Level-2 Residuals, Texas

<u>Spatial Dependence Structure</u>	<u>Lagrange Multiplier</u>
Error	9.43**
Lag	15.46***
Error (robust) ^a	3.33 [†]
Lag (robust) ^a	9.36**

^a“Robust” indicates that the associated Lagrange Multiplier estimate is robust to the other form of spatial dependence.

*
p < 0.05,

**
p < 0.01,

p < 0.001

Table 4

Spatial HLM Results, Texas

	<u>Odds Ratio</u>
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)	2.46***

Notes: Individual coefficients are unchanged from baseline.

*
p < 0.05,

**
p < 0.01,

p < 0.001