Published in final edited form as:

Med Care Res Rev. 2008 June ; 65(3): 315–337. doi:10.1177/1077558707312501.

How Sensitive Are Multilevel Regression Findings to Defined Area of Context? A Case Study of Mammography Use in California

Lee R. Mobley, Tzy-Mey (May) Kuo, and Linda Andrews

RTI International, Research Triangle Park, NC

Abstract

The authors develop a hybrid model of health care use that blends features of the traditional Aday–Andersen behavioral model with the socioecological modeling perspective. They use the model to conceptualize the various levels of influence expected from socioecological variables in individuals' mammography use decisions, build contextual variables from fine-grained data into four different types of geographic areas, and then use two- and three-level modeling of personal and area-level contextual factors to explain observed behavior. The central focus is on whether differentiating the conceptualized levels of influence seems to materially affect regression findings. The test could conceivably be confounded by the modifiable areal unit problem, but little evidence for this is found. Findings for California women suggest that distinctions do matter in how the levels of influence are defined for local neighborhood contextual factors. Studies using only county-level contextual factors will miss some meaningful associations related to interpersonal/proximate-level factors.

Keywords

access to care; mammography use; modifiable areal unit problem; multilevel model; socioecological model; breast cancer screening

There are many different ways to combine microdata into variables describing small areas, and how this is done may affect statistical properties of the variables (Ali et al., 2005; Tatalovich, Wilson, Milam, Jerrett, & McConnell, 2006). Greenland (2002) and Pickett and Pearl (2000) discussed what this might mean for multilevel analyses. Empirical small area analyses (with areal-level aggregates for both the dependent and the explanatory variables) have found inconsistencies across models using different areal aggregations, which has been named the modifiable areal unit problem, or MAUP (Cockings & Martin, 2005; Fotheringham & Wong, 1991). To our knowledge, there have been no published studies with empirical assessments of MAUP in multilevel regression models, which is a major contribution of this article.

To test for MAUP in multilevel models requires careful conceptualization of the relevant zones of influence for contextual variables (i.e., the appropriate level of aggregation for use in the analysis). A strong conceptual model is needed to guide the assessment because coefficient estimates can change for other reasons besides MAUP. These include differences in the relevance of the zones of influence defined (Pickett & Pearl, 2000), differences in statistical properties of the constructed variables because of MAUP (Greenland, 2002), or omission of important variables (Dietz, 2002).

Please address correspondence to Lee R. Mobley, RTI International, 3040 Cornwallis Road, 275 Cox Building, Research Triangle Park, NC 27709-2194; email: E-mail: lmobley@rti.org.

Authors' Note: This article, submitted to *Medical Research and Review* on October 31, 2007, was revised and accepted for publication on November 19, 2007.

To identify all important variables and their relevant zones of influence, we developed a hybrid conceptual model that blends the best features of traditional behavioral models of access and use with socioecological models of health outcomes. This model situates the individual decision maker, distinguished by enabling–predisposing need characteristics, into an ecological context that has different zones of influence (fundamental–distal, intermediate–community, and interpersonal–proximate). A finding that the MAUP does not seem to matter would allow interpretation of observed differences between models using larger versus smaller areas as being a result of variations in the zone of influence inherent in the modeling. Using several different areal definitions, we assessed whether model coefficients change with areal units of aggregation and interpret the findings on the basis of our conceptual model.

New Contribution

Analyses of the use of breast cancer screening have looked at a variety of factors, but they usually include only distance to provider as a measure of the geospatial environment. Only recently have studies included socioecological factors such as marital status or living in female-headed households, urbanization, or medically underserved area designation (Gumpertz, Pickle, Miller, & Bell, 2006; Litaker & Tomolo, 2007). With funding from the National Cancer Institute, we have developed several new geospatial variables at different geographic levels, and these are freely available to other researchers. To assess the importance of these new variables on mammography use requires multilevel modeling, combining personal and area-level factors. Perhaps more important, it requires a determination of what the local area should be, that is, how it should be defined in geographic space. In this work, we built a conceptual model of the various zones of influence for contextual variables and then determined whether findings differ across models with different modeling of areal units. This work informs the socioecological modeling literature and provides useful socioecological variables for public use.

Background

Modifiable Areal Unit Problem

Areal units are geographic entities such as counties or states. Zonal partitioning schemes are decisions about how to draw the boundaries for areal units. The MAUP emerges when data developed as contextual factors for various areal units have different statistical properties. When the underlying microdata values are not uniformly distributed across the landscape, merging smaller units into larger areas using different zonal partitioning schemes is the same as smoothing (averaging) across different combinations of spatial neighbors (Wong, 1996). When these zonal schemes overlap (are not nested), the different combinations of spatial neighbors can yield constructed variables with quite different statistical properties (Ali et al., 2005; Greenland, 2002; Tatalovich et al., 2006). Both signs and significance levels of regression coefficient estimates can be affected, and considerable variability has been found across small-area studies examining essentially the same phenomenon (Cockings & Martin, 2005; Fotheringham & Wong, 1991; Waller & Gotway, 2004).

Potential for MAUP in Our Data

The four areal units used in our California analysis are the county ($n = 57$), the medical service study area (MSSA; $n = 519$), the primary care service area (PCSA; $n = 333$), and the U.S. Census ZIP code tabulation area (ZCTA; $n = 1,450$).¹ Table 1 gives definitions of these areal units and information about nesting and non-nesting of these with other areal units. Areas that

¹MSSAs were defined by the California Office of Statewide Health Planning and Development (2005); ZCTAs were defined by the U.S. Census Bureau (2000a).

do not nest have overlapping boundaries, such as ZCTAs and counties. Areas that do nest are contained within or perfectly contain one another, such as MSSAs and counties. For example, Los Angeles County has 295 ZCTAs, 44 PCSAs, and 98 MSSAs; PCSA or ZCTA and MSSA boundaries cross, while ZCTAs are nested inside PCSAs.

Both MSSAs and PCSAs were defined (by the California Office of Statewide Health Planning and Development, 2005, and the Health Resources and Services Administration, n.d., respectively) on the basis of economic principles but for different populations. MSSAs partition counties into smaller areas on the basis of the health resource needs of the entire population and degree of rurality. MSSA markets are defined for the entire population and nest within the county, dividing them into subregions that are classified as urban, rural, or frontier (Christman, 2004). By contrast, PCSAs are physician market areas derived from observed flows of elderly patients to primary care providers (Goodman et al., 2003). Both the PCSA and MSSA areas are much smaller than counties, which in California are so large that both urban and rural areas are contained within single counties. Thus, PCSAs and MSSAs do a much better job than counties in separating rural from nonrural areas. The ZCTA is the smallest area we are able to study because of privacy limitations on the individual-level data.

The overlapping PCSA and MSSA boundaries and differences in partitioning of the underlying populations can yield MAUP. This is illustrated in Figures 1 and 2, which show the Los Angeles County, California, area, picked as an example to show how the MAUP problem can arise as a result of underlying heterogeneity in the microdata. Figure 1 shows the census of elderly women in Los Angeles County partitioned into 98 MSSAs, which nest completely inside and cover all spaces within the county. Figure 2 shows the census of elderly women in the Los Angeles County area partitioned into 44 PCSAs, several of which cross the county boundary. The PCSA map (Figure 2) is based on 574,592 elderly women residing in 44 PCSAs centered in (but not fully contained within) Los Angeles County. The MSSA map (Figure 1) is based on 541,915 women in 98 MSSAs completely inside Los Angeles County. To illustrate the potential for the MAUP, the counts of elderly women (variable) in these MSSA and PCSA areas are represented in the maps using standard deviations from their means. Contrasting patterns in the two maps show the different statistical properties of this same variable when partitioned into MSSAs versus PCSAs.

There are other differences in the MSSA and PCSA areal coverages. In Figure 2, there are some holes (missing areas) evident in the PCSAs; these are areas within Los Angeles County built up from alphanumeric ZCTAs, which cover water bodies or uninhabited lands and have no census data attached to them. More generally, the PCSA areal coverage has holes in places that are not inhabited, while the MSSA coverage includes all land areas and exhibits no holes. (Note that these holes coincide with MSSAs where the counts of elderly women are sparse, more than 2 standard deviations below the MSSA mean.)

The Hybrid Conceptual Model

The behavioral model of use developed by Aday and Andersen (A-A; 1974) is frequently used in describing factors associated with health care access and use, but these studies rarely include many contextual factors (Andersen, 1995; Phillips, Morrison, Andersen, & Aday, 1998). By contrast, models of health outcomes commonly use contextual factors at multiple levels within a socioecological framework (Booth et al., 2001; Northridge, Sclar, & Biswas, 2003; Schulz et al., 2005; Smedley & Syme, 2000).

Although the idea that the community influences human behavior has been around for some time (Bronfenbrenner, 1979), explicit consideration of the relevant zone of influence for ecological variables has only recently begun to appear in the socioecological literature (Booth et al., 2001; Northridge et al., 2003; Pickett & Pearl, 2000; Sampson, Morenoff, & Gannon-

Rowley, 2002; Schulz et al., 2005; Smedley & Syme, 2000). Pickett and Pearl (2000) argued that findings may vary across studies using different zonal schemes precisely because the so-defined areal units presume very different zones of influence for the contextual variables, when these distinctions truly matter.

In this article, we blend the health behavioral and socioecological perspectives in our unique hybrid spatial interactions model. We begin with the model of health care access and use developed by Khan and Bhardwaj (1994), which defines the use outcome as a process involving the interplay between the characteristics of the health care service system, potential users, and the social and physical features of places. This model provides a distinctly spatial view of human interaction with the environment and other structural and social aspects of the community. Some limitations of the Khan and Bhardwaj (K-B) model are that it does not consider the levels of influence that are central to the mainstream socioecological perspective or differentiate personal characteristics into the enabling–predisposing need constructs in the A-A behavioral health model. These constructs are useful when considering multilevel interactions between people and environmental characteristics along the pathways to health care use (Litaker, Koroukian, & Love, 2005).

We adapt the K-B model to create a unique hybrid of the K-B and A-A models and the socioecological perspective, as shown in Figure 3. The white “health care system” and “individual characteristics” boxes reflect constructs in the K-B model, which are filtered through contextual factors in a separate barrier–facilitators box (not shown), determining observed use. This K-B barriers–facilitators box is represented in Figure 3 by a series of nested constructs: the large box surrounding the entire figure (fundamental–distal factors) and the two smaller nested ovals within it (intermediate–community and interpersonal–proximate factors). The two smaller ovals partition the K-B contextual variables into three different levels of influence, consistent with the traditional socioecological model. The hybrid model thus views the individual as making use choices (the final outcome) in a market context that has differentiated levels of influence for different classes of contextual variables.

The conceptual model guides the selection of variables to be included in the analysis and suggests the appropriate level of aggregation for contextual variables through the classification of zones of influence. The variables chosen for this work, where they fit in terms of the concept, and their sources are described in Table 2, which is divided into three sections: individual characteristics (categorized into enabling, predisposing, and need categories), interpersonal–proximate factors, and intermediate–community factors. Each variable is described as to where it fits in the conceptual model. The interpersonal–proximate factors are measured at three different levels of aggregation reflecting local communities (ZCTA, PCSA, and MSSA) so that we can assess whether these different definitions matter empirically. The intermediate–community factors are measured at the county level, which is larger than the local community areas. The fundamental–distal factors category is not represented in the analysis because we view these as state- or regional-level variables.

Individual characteristics—An important enabling characteristic is type of health coverage. All women in the sample had at least the traditional fee-for-service (FFS) Medicare insurance, which allows free choice of provider and mammography facility. Persons without Part B insurance coverage, which covers outpatient services but requires payment of an additional premium, were excluded from the analysis because no mammography claims are available for these people. The same is true of women with Medicare health maintenance organization (HMO) insurance. If a woman had Medicare HMO coverage in the 2 years before the study period (2002–2003), then an indicator was created to control for the fact that a recent mammogram might have been obtained under the HMO, which could lower the probability of use in 2002–2003. Some women have additional coverage because of personal disability, which

makes them dually eligible for both Medicare and Medicaid, whereas others are dually eligible on the basis of low income or other considerations. Although there are additional resources, we expect that disability and dual eligibility status are disabling characteristics because physical limitations and poverty present additional burdens to care-seeking behavior.

Shorter distance to the closest mammography facility is seen as an enabling characteristic. Another characteristic not seen in previous research is recent address change; we expected moving to be disruptive and a disabling characteristic. Predisposing factors included in the model are age and race or ethnicity. We include cancer diagnosis and use of flu shots as indicators of need. Persons with a previous cancer diagnosis are more likely to experience another cancer, and mammography is used in the course of treatment as a diagnostic. Those using flu shots are considered to have stronger health-seeking behavior. We do not have information on personal income, marital status, employment status, or education.

Interpersonal–proximate factors—These include local community or neighborhood characteristics that affect a person’s perception of risk or information about health care, through interactions with neighbors that shape opinions and beliefs. Communities may also provide support—both physical and psychological—for health-seeking behaviors. While residential segregation is often viewed as a harmful fundamental factor—because it can influence the distribution of wealth, opportunity, and political influence toward the majority—in the local neighborhood, residential segregation may affect social integration and support. We use Massey and Denton’s (1988) isolation index² as our residential segregation measure, defined separately for each race or ethnicity relative to Whites and recalculated for each type of areal unit. We expect that at the interpersonal level, the index may have positive impacts for some groups and negative impacts for others because residential segregation has been associated with both better and worse health outcomes, depending on the race or ethnicity of the study population (Mobley, Root, Finkelstein, Khavjou, & Will, 2006; Palloni & Arias, 2004; Schulz, Williams, Israel, & Lempert, 2002; Williams & Collins, 2001).

Two variables were included to reflect social or cultural cohesion: the proportion of community members who have recently immigrated into the United States and the proportion of elderly community members with little or no English language ability. Both of these variables were expected to reduce cohesion and the probability of mammography use. Several stressor variables were included for each woman’s local community: commuter intensity, elderly women in poverty, and elderly women living alone. Interdriver courtesy, which in our real-life experience decreases in communities with high commuter intensity, is expected to affect the difficulty experienced by elderly persons who drive or for their caregivers who drive them. Areas with greater commuter intensity have been found to exhibit lower access to preventive care services among elderly persons (Mobley, Root, Anselin, Garcia, & Koschinsky, 2006); thus, we expected that commuter intensity would reduce mammography use. We expected that areas with higher proportions of elderly women living in poverty or living alone would exhibit

²The isolation index is an exposure measure that represents the propensity for a minority member to be exposed only to other minority members. It is actually a sort of probability and ranges in value from 0 to 1. Higher values represented greater isolation or segregation. For the county-level measure, the isolation index was computed as the minority-weighted average of each tract’s minority population:

$$P^{jm} = \sum_{i=1}^N \frac{x_i}{X} \frac{x_i}{t_i}$$

where j represents the county unit, m is the minority group, i represents the tract unit, N is the number of tracts within the county, x_i is the minority population in tract i , t_i is the total population (all races) in tract i , and X is the total minority population in the county.

lower mammography use rates because of lower social and material support and that women living in such areas would exhibit lower probability of use.

Intermediate–community factors—At the wider community level are social context and physical environment factors that are affected by fundamental–distal resources that shape the infrastructure supporting community life (Schulz et al., 2005). Important among these are characteristics of the health care system such as physician shortage, facility density and proximity, and managed care climate. Managed care penetration in an area can change the way in which medicine is practiced, with spillover effects on FFS Medicare patients (Baker, 2003), so we expected that women living in areas with greater managed care penetration—although not in managed care plans themselves—might exhibit different probabilities of mammography use. Use would be higher if area attitudes regarding prevention were enhanced, but lower if crowding at facilities was subsequently greater, so we had no a priori expectation regarding the sign of the HMO spillover effect.

Availability of medical oncologists and nurses was expected to increase probability of mammography use. Women living in areas with primary care physician shortages were expected to have lower probability of use. Primary care physician shortage is indicated using Health Resources and Services Administration's measure at the county level. An alternative measure of physician availability is the ratio of international medical graduates (IMGs)—physicians of foreign origin who train in the United States—to native U.S.-born physicians. One study has found that IMGs have disproportionately located in U.S. counties of greatest need compared with native medical graduates (Mick, Lee, & Wodchis, 2000), which reflects successful efforts through the J-1 visa waiver program that encourages this ("Role of International Medical Graduates," 2003). Another study found that poor rural places with higher ratios of IMGs to native physicians had improved access to preventive care services for elderly persons (Mobley, Root, Anselin, et al., 2006). However, IMGs tend to settle in both urban and rural enclaves ("Role of International Medical Graduates," 2003), so the expected impact of IMG presence on a woman's mammography use is uncertain a priori.

Study Population, Data, and Methods

We used 2000–2003 data from the Surveillance Epidemiology and End Results–Medicare linkage of cancer registry records, a Medicare 5% reference sample, and their medical claims. We examined mammography use in 2002–2003 among women at least 65 years of age with traditional FFS Medicare insurance (both Parts A and B) who resided in California. The geographic areas in the state of California span the urban–rural continuum and provide good coverage of women of various races and ethnic backgrounds. There is also good density of mammography service locations, which is typical of most of the United States (U.S. General Accounting Office, 2002). More than half the women in the sample lived in the same ZIP code as a mammography facility, and only 3% were located in a ZIP code more than 10 miles from a facility. National guidelines recommend annual mammography screening for women older than age 40; Medicare has covered this service since 1998, but fewer than half of the FFS sample women used mammography over this 2-year interval.

Empirical Model

A multilevel binary probit model was used to examine how the contextual factors affect screening uptake among elderly California women. The dependent variable in the probit model is a binary variable indicating whether a woman received at least one mammogram during the years 2002 and 2003. The predicted value from the regression is a Z score, which follows the standard normal density distribution. The estimated parameters were converted into marginal probabilities when presenting the findings, so coefficient estimates can be interpreted as

changes in probability of use associated with a small (1-unit) change in the covariate. We used SAS general estimating equations to correct the standard errors for biases stemming from redundancies in the contextual variables within areal units.

The empirical model includes person- and area-level variables, and four versions of the model were estimated, each with a different areal unit for the interpersonal–proximate contextual variables (ZCTA, PCSA, MSSA, and county). The first three models (ZCTA, PCSA, and MSSA) are three-level models with county-level variables at the intermediate–community level. The fourth model is a two-level model that uses the county as the areal unit for all contextual variables, to enable answering our study questions about levels of influence. The variables and their sources are described in Table 2; Tables 3 and 4 contain sample statistics.

Because physician markets based on elderly person flows to doctors (the basis for PCSAs) and resource needs (the basis for MSSAs) are both based on economic factors, they are both relevant market definitions from an economic perspective and thus useful for small-area analyses. For the current analysis, however, they are used to approximate the zones of influence for various interpersonal–proximate contextual factors, which reflect small communities or neighborhoods that are not necessarily the same as health market areas. However, both areal units differentiate urban and rural areas as separate places, in contrast to counties, which include both in their borders. Because both of these area-level contextual variables account for urban intensity and are defined over areal units of similar size, we expect their levels of influence to be fairly similar but different from their county-level counterparts. However, the overlapping PCSA and MSSA boundaries and differences in partitioning of the underlying populations can yield MAUP (i.e., derived variables with different statistical properties); thus, dissimilar estimation results might emerge because of differences in the ways that the underlying heterogeneous data are partitioned.

ZCTAs are the smallest areal units that we can define, given privacy restrictions within the data. ZCTAs are zones used for mail delivery and may be more representative of small neighborhoods than the MSSA or PCSA measures, although they are most likely not ideal neighborhood definitions. ZCTAs nest inside PCSAs but cross MSSA borders, so we might expect more MAUP in the comparison of ZCTA and MSSA findings. Similarly, MSSAs nest inside counties, so we might expect less MAUP in comparisons of MSSAs and counties than in comparisons of MSSAs and PCSAs. In the empirical analysis, we answer the following research questions:

Question 1: Are the empirical findings similar across the three three-level models?

Question 2: If the three-level model findings are similar (thus providing no evidence of MAUP effects), are they also similar to the two-level (county) model findings?

If the answer to Question 1 is “yes,” then despite the crossing boundaries and different partitioning of underlying data when creating the areal-level data for the ZCTAs, PCSAs, and MSSAs, no MAUP is evident. If no MAUP is evident across these three areal units, then we would expect none with the fourth (county). Therefore, we would expect the answer to Question 2 to be “yes,” unless there are true zone-of-influence differences, since the county is much larger than the other three areal units. With this logic, the answers to these questions will help us determine whether there are apparent differences in findings deriving from the assumed zones of influence imposed by the areal units used in the models.

Findings and Discussion

The regression results are provided in Table 5 for each of the four models.

Individual Characteristics

The impacts of person-level variables are essentially the same across the four models, with the exception of the race or ethnicity variables. For Asians, African Americans, American Indian/Alaska Natives, and other, the findings are comparable across models; being of these races or ethnicities reduces the probability of mammography use, relative to Whites (the reference group). Hispanic ethnicity does not significantly affect use in any model except the county-level model, where there is a weakly significant negative effect. MAUP or zone-of-influence effects were not expected on the individual-level variables, and findings confirm this.

Interpersonal-Proximate Factors

The isolation index reflects residential segregation of minority groups and is (mostly) significant and negative for both African American and Hispanic variables. These are interpreted as lower probability of mammography use for women living in communities where there is higher segregation of African Americans or Hispanics. This finding is consistent across all four models, but significance levels are lower in the county-level model. The consistency in sign across models suggests that there is no MAUP effect, while the reduced significance at the county level suggests that there are perhaps important zone-of-influence effects. More specifically, these isolation variables seem to influence outcomes at areal units much smaller than the county.

Women living in communities with greater segregation of American Indian/Alaska Native populations seem more likely to use mammography, which is opposite the finding for women in communities with higher segregation of African Americans or Hispanics. This suggests that there is social support in segregated American Indian/Alaska Native communities that promotes use of mammography. For women living in these communities, the estimated association is large and positive in three models but not in the MSSA model, whereas both the ZCTA (smallest area) and county (largest area) models show significant positive coefficients. These findings suggest that there are MAUP effects causing inconsistent findings across the MSSA and other models for this variable, but this evidence suggests no distinctions between different zones of influence from small to large communities.

Regarding the MAUP effects, Native American populations in California tend to be most segregated in the isolated rural, frontier areas of MSSAs. These areas often coincide with the rural holes (noted earlier) in the PCSA coverage, which means that MSSAs and PCSAs will vary considerably in how they carve up the underlying populations used in calculating the American Indian/Alaska Native isolation index. No other racial or ethnic group in California has high segregation in rural areas, which may explain why we see MAUP for this variable and not others.

Women living in communities with greater proportions of those with long commutes (traveling 60 min or more each way to work) are significantly less likely to use mammography across all models. This is consistent with previous findings in small-area analysis of PCSA markets regarding commuter intensity and use of preventive services by elderly persons (Mobley, Root, Anselin, et al., 2006). Two other stressors expected to have negative associations are living in communities with greater proportions of elderly persons in poverty or communities with greater proportions of elderly women living alone. The elderly persons in poverty and women living alone coefficients are quite comparable across the models, and they are significant in all but the county model. This finding suggests that there is no MAUP, but there may be important zone-of-influence effects. These stressors seem more appropriately defined at the smaller areal units, which is consistent with our conceptual model. However, we did not expect to find a positive impact on a woman's probability of use from living in communities where a greater proportion of elderly women live alone. This finding warrants further study.

Intermediate–Community Factors

The community-level factors were entered at the county level in all models, following the zones of influence identified in our conceptual model, so we did not expect to see any MAUP effects across models. Nurses per capita, violent crime rates, and land use mix are consistent, significant predictors across all four models. Physician variables do not have significant impacts, and managed care climate is only weakly significant. It is not surprising that elderly women living in areas with higher crime rates are less likely to use mammography. The land use mix variable is larger when communities are more diverse and urban, and living in a place with a more urban aspect seems to discourage mammography use. Managed care penetration seems to have a small positive effect, suggesting beneficial spillovers in preventive care behaviors from managed care presence in the community, but the evidence is statistically weak.

Limitations

Further elaboration of the model could include explicit interaction effects between the contextual factors at different levels and the enabling characteristics of individuals. We exclude these interactions for simplicity but include a multitude of factors and characteristics to make the model as general as possible. When applying the model to a particular problem, one caveat is that all items are not equally influential in modeling different aspects of access or use for different populations or services, and interaction effects may vary with the problem under study. It is important to consider carefully all of the relevant variables that should be included in the model because omitted variables bias is another source of inconsistency in parameter estimates across studies using different zonal schemes (Dietz, 2002). Thus, we may have erred, providing fruitful opportunities for future research.

In the single time interval examined here, this cross-sectional analysis is limited in its ability to infer causal effects; thus, the relationships discovered are merely statistical associations. The contextual variables all predate the outcome variable (Table 2) so they are predetermined; even so, there may be simultaneity within the probability-generating spatial interaction. Another limitation may be the ability to generalize the policy findings to other areas or states. Finally, we were not able to examine areas smaller than ZCTAs for our assessment of MAUP. However, it is unlikely that data would ever be available at finer areal units because of privacy concerns.

Summary and Conclusions

MAUP effects are considered important when they are present in the data because researchers and policymakers would have the opportunity to pick and choose among various contradictory results to find some most consistent with a professed theory or agenda. We have seen no empirical assessment of MAUP effects in multilevel models in the literature, which is this article's major contribution. This assessment is important because researchers would like to know which size of area is relevant for assessing contextual factor impacts, which may have different zones of influence across study variables.

We developed a conceptual model with hypothesized zones of influence for the contextual variables, following the socioecological literature. We estimated four multilevel models of mammography use decisions and used results to assess MAUP and to study zone-of-influence effects. The first three models used different definitions of local areas (ZCTA, PCSA, and MSSA) for interpersonal contextual variables and larger county areas for intermediate–community-level variables. In these three-level models, the local and community areas overlap and have features conducive to generating MAUP. The fourth (county) model has only two levels, using the county area to represent both interpersonal–local and intermediate–community variables—rather than differentiating local and wider community levels of influence. With this setup, we first look for evidence of MAUP. Finding none, we can then

interpret differences across the three- and two-level models as suggesting that zone-of-influence effects are important.

We find very little evidence of MAUP, despite the fact that features of the underlying microdata and areal units could conceivably generate it. There is a suggestion of MAUP for one variable: the local area segregation index for American Indian/Alaska Native populations. This variable exhibits characteristics that are quite conducive to MAUP, namely having higher values in more sparsely settled areas, so this finding is not surprising. To assess zone-of-influence effects, we compared findings across the three- and two-level models to assess whether coefficient estimates lose or gain significance when defined at a larger or smaller scale. We find that several of the interpersonal–proximate-level factors that are significant in all of the three-level models lose significance in the fourth, two-level model. The importance of this finding is that researchers using county-level data to measure these contextual factors would not have found any significant associations, whereas others using more local definitions would have. These factors include residential segregation of African Americans and Hispanics and measures of elderly poverty and isolated living arrangements in the local community.

The level of aggregation for multilevel socioecological models should be carefully considered and defined appropriately. Data developed for this work under National Cancer Institute grant funding are available to other researchers. We hope that this availability will allow that these findings can be tested and validated by others in this fruitful area of research.

Acknowledgements

The project described in this article was supported by Grant Number 1 R21 CA117869 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

References

- Aday LA, Andersen R. A framework for the study of access to medical care. *Health Services Research* 1974;9:208–220. [PubMed: 4436074]
- Ali, M.; Park, J.; Thiem, V.; Canh, D.; Emch, M.; Clemens, J. Neighborhood size and local geographic variation of health and social determinants. *International Journal of Health Geographics*. 2005 Jun. Retrieved from <http://www.ij-healthgeographics.com/content/4/1/12>
- Andersen RM. Revisiting the behavioral model and access to medical care: Does it matter? *Journal of Health and Social Behavior* 1995;3:1–10. [PubMed: 7738325]
- Baker LC. Managed care spillover effects. *Annual Review of Public Health* 2003;24:435–456.
- Booth SL, Sallis JF, Ritenbaugh C, Hill JO, Birch LL, Frank LD, et al. Environmental and societal factors affect food choice and physical activity: Rationale, influences, and leverage points. *Nutrition Review* 2001;59(3 Pt 2):S21–S39.
- Bronfenbrenner, U. *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press; 1979.
- California Office of Statewide Health Planning and Development. Medical service study areas. 2005. Retrieved December 7, 2007, from <http://gis.ca.gov/catalog/BrowseRecord.epl?id=23784>
- Christman, JS. Working paper. University of Southern California: School of Policy, Planning, and Development; 2004 Sep. Urban and rural designations: The impact on rural healthcare in California. Retrieved December 7, 2007, from http://www.ruralhealth.ca.gov/pdf/Urban_Rural_issue_Paper.pdf
- Cockings S, Martin D. Zone design for environment and health studies using pre-aggregated data. *Social Science and Medicine* 2005;60:2729–2742. [PubMed: 15820583]
- Dietz R. The estimation of neighborhood effects in the social sciences: An interdisciplinary approach. *Social Science Research* 2002;31:539–575.
- Fotheringham A, Wong D. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning* 1991;23:1025–1044.

- Goodman DC, Mick SS, Bott D, Stukel T, Chang C, Marth N, et al. Primary care service areas: A new tool for the evaluation of primary care services. *Health Services Research* 2003;38:287–310. [PubMed: 12650392]
- Greenland S. A review of multilevel theory for ecologic analyses. *Statistics in Medicine* 2002;21:389–395. [PubMed: 11813225]
- Gumpertz ML, Pickle LW, Miller BA, Bell BS. Geographic patterns of advanced breast cancer in Los Angeles: Associations with biological and sociodemographic factors (United States). *Cancer Causes & Control* 2006;17:325–339. [PubMed: 16489540]
- Health Resources and Services Administration. Primary care service areas. nd. Retrieved December 7, 2007, from <http://datawarehouse.hrsa.gov/pcsa.htm>
- Khan AA, Bhardwaj SM. Access to health care: A conceptual framework and its relevance to health care planning. *Evaluation & the Health Professions* 1994;17:60–76.
- Litaker D, Koroukian SM, Love TE. Context and healthcare access: Looking beyond the individual. *Medical Care* 2005;43:531–540. [PubMed: 15908847]
- Litaker D, Tomolo A. Association of contextual factors and breast cancer screening: Finding new targets to promote early detection. *Journal of Women's Health* 2007;16:36–45.
- Massey DS, Denton NA. The dimensions of residential segregation. *Social Forces* 1988;67:281–315.
- Mick S, Lee S, Wodchis W. Variations in geographical distribution of foreign and domestically trained physicians in the United States: “Safety nets” or “surplus exacerbation”? *Social Science & Medicine* 2000;50:185–202. [PubMed: 10619689]
- Mobley, LR.; Root, ED.; Anselin, L.; Garcia, NL.; Koschinsky, J. Spatial analysis of elderly access to primary care services. *International Journal of Health Geographics*. 2006 May. Retrieved from <http://www.ij-healthgeographics.com/content/5/1/19>
- Mobley LR, Root E, Finkelstein E, Khavjou O, Will J. Relationships between the built environment and other contextual factors, obesity, and cardiac risk. *American Journal of Preventive Medicine* 2006;30:327–332. [PubMed: 16530620]
- Northridge M, Sclar E, Biswas P. Sorting out the connections between the built environment and health: A conceptual framework for navigating pathways and planning healthy cities. *Journal of Urban Health* 2003;80:556–568. [PubMed: 14709705]
- Palloni A, Arias E. Paradox lost: Explaining the Hispanic adult mortality advantage. *Demography* 2004;41:385–415. [PubMed: 15461007]
- Phillips KA, Morrison KR, Andersen R, Aday LA. Understanding the context of healthcare utilization: Assessing environmental and provider-related variables in the behavioral model of utilization. *Health Services Research* 1998;33(3 Pt 1):571–596. [PubMed: 9685123]
- Pickett KE, Pearl M. Multilevel analyses of neighborhood sociodemographic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health* 2000;54:111–122.
- Role of international medical graduates in America's small rural “critical access” hospitals. *CAH/FLEX National Tracking Project: Findings From the Field* 2003;3(1)
- Sampson R, Morenoff J, Gannon-Rowley T. Assessing “neighborhood effects”: Social processes and new directions for research. *Annual Review of Sociology* 2002;28:443–478.
- Schulz AJ, Kannan S, Dvonch JT, Israel BA, Allen A 3rd, James SA, et al. Social and physical environments and disparities in risk for cardiovascular disease: The healthy environments partnership conceptual model. *Environmental Health Perspectives* 2005;113:1817–1825. [PubMed: 16330371]
- Schulz AJ, Williams DR, Israel BA, Lempert LB. Racial and spatial relations as fundamental determinants of health in Detroit. *Milbank Quarterly* 2002;80:677–707. [PubMed: 12532644]
- Smedley, BD.; Syme, SL., editors. *Promoting health: Strategies from social and behavioral research*. Washington, DC: National Academies Press; 2000.
- Tatalovich, Z.; Wilson, J.; Milam, J.; Jerrett, M.; McConnell, R. Competing definitions of contextual environments. *International Journal of Health Geographics*. 2006 Dec. Retrieved from <http://www.ij-healthgeographics.com/content/5/1/55>
- U.S. Census Bureau. Appendix A: Census 2000 geographic terms and concepts. 2000a. Retrieved December 7, 2007, from <http://www.census.gov/geo/www/tiger/glossry2.pdf>

- U.S. Census Bureau. ZIP code tabulation areas (ZCTAs). 2000b. Retrieved December 7, 2007, from <http://www.census.gov/geo/ZCTA/zcta.html>
- U.S. General Accounting Office. Mammography: Capacity generally exists to deliver services (Pub. No. GAO-02-532). Washington, DC: Author; 2002. Available from <http://www.gao.gov/new.items/d02532.pdf>
- Waller, LA.; Gotway, CA. Applied spatial statistics for public health data. Hoboken, NJ: John Wiley; 2004.
- Williams DR, Collins C. Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports* 2001;116:404–416. [PubMed: 12042604]
- Wong, D. Aggregation effects in geo-referenced data. In: Arlinhaus, S., editor. *Practical handbook of spatial statistics*. Boca Raton, FL: CRC Press; 1996. p. 83-106.

Figure 1.
Los Angeles County: Counts of Total Elderly Female Population in 98 Medical Service Study Areas (MSSAs)

Figure 2.
Los Angeles County, Counts of Total Elderly Female Population in 44 Primary Care Service Areas (PCSAs)

Figure 3.
Spatial Interaction Model of Mammography Use

Table 1

Various Areal Units, With Nesting (Hierarchical) or Overlapping Boundaries

Areal Unit	Definition	Nesting Units	Overlapping (Non-Nesting) Units
Census tract ^a	Census tracts are small, relatively permanent statistical subdivisions of a county, with between 1,500 and 8,000 people, and an optimum size of 4,000 people. Counties with fewer than 1,500 people have a single census tract.	County, MSSA	ZCTAs, PCSAs
County ^a	Counties are the primary legal divisions of most states; in Louisiana, these are known as parishes. Alaska has boroughs, and some cities in Virginia are classified as counties.	Census tract, MSSA	ZCTAs, PCSAs
ZCTA ^a	ZCTAs (ZIP code tabulation areas) are a statistical geographic entity that approximates the delivery area for a U.S. Postal Service five-digit ZIP code (or three-digit in rural areas) or collection of ZIP codes.	PCSAs	Counties, MSSAs
PCSA	PCSAs are primary care service areas defined by Dartmouth College researchers for Health Resources and Services Administration, built up from Medicare fee-for-service patients' flows to primary care physician, using ZIP code address of person (Goodman et al., 2003).	ZCTAs	Counties, MSSAs
MSSA	MSSAs are medical service study areas, defined by the state of California to break up counties into smaller areas reflecting urban, rural, and frontier (very remote) regions (Christman, 2004).	Census tracts, counties	PCSAs, ZCTAs

^aSource: U.S. Census Bureau (2000a).

Table 2
Variables Chosen for Analysis, Their Contextual Relevance, and Sources

Variable	Source
Individual characteristics of sample women	
Enabling	
Personal disability as original reason for Medicare entitlement	Developed from linked California Surveillance and Epidemiology and End Results cancer registry and Medicare data, provided by the National Cancer Institute, 2002–2003
Moved to a new residence in 2002–2003	
Months enrolled in a Medicare health maintenance organization anytime in 2000–2001	
Months with dual eligibility-where entitled to both Medicare and Medicaid, 2002–2003	
Distance to closest mammography facility	
Predisposing	
Age in 2002	
Race or ethnicity (White, African American, Asian, Hispanic, American Indian/Alaska Native, other)	
Need	
Had cancer diagnosis on file in 2002–2003	
Obtained flu shot from doctor in 2002–2003	
Interpersonal-proximate factors (ZCTA, MSSA, PCSA)	
Social integration and support: Isolation index describing segregation by race or ethnicity (White, African American, Asian, Hispanic, American Indian/Alaska Native, other)	Developed from U.S. Census (2000b) microdata at tract and ZCTA levels
Stressor, driver courtesy: Commuter intensity reflecting the proportion of workforce in each woman's residential area commuting 60 min or more each way to work	
Social or cultural cohesion:	
Proportion of population in each woman's residential area immigrating into the United States between 1995 and 2000	
Proportion of elderly persons in each woman's residential area with little or no English language ability	
Stressor, living conditions	
Proportion of female elderly persons in each woman's residential area living alone	
Proportion of elderly persons in each woman's residential area living below the federal poverty level	
Intermediate-community factors (county)	
Health care system	Surveillance Epidemiology and End Results-Medicare linked data and U.S. census
No. mammography facilities per 1,000 elderly women in each woman's residential area, 2002–2003	
No. oncologists (or nurses) per 1,000 elderly persons, 2001 (2000)	Area resource file
Ratio of counts of international medical graduates to U.S.-born physicians (generalists and specialists), 2000–2001	Health Resources and Services Administration geospatial data warehouse (http://datawarehouse.hrsa.gov/)
County designated as a primary care provider shortage area in whole or part of county, 2002: http://bhpr.hrsa.gov/shortage/hpsacritpcm.htm	Area resource file from Health Resources and Services Administration
Managed care penetration, defined as proportion of the county population insured by health maintenance organizations, 2002–2003	InterStudy
Social context: Number of violent crimes known to police per 1,000 population	U.S. Department of Justice and Federal Bureau of Investigation
Physical environment: Index of Land Use Mix	National land use/land cover data, U.S. Geological Survey

Variable	Source
----------	--------

Note: MSSA = medical service study area; PCSA = primary care service area; ZCTA = ZIP code tabulation area.

Table 3

Sample Statistics for Person-Level Variables: Means (and Standard Deviations) or Proportions, Expressed as Percentages

Variable	Total (<i>n</i> = 70,129)	Women With Mammography (<i>n</i> = 41,376)	Women Without Mammography (<i>n</i> = 28,753)
Age as of 2002	76.5 (7.2)	75.3 (6.4)	78.3 (7.9)
% with breast cancer	32.58	45.02	14.65
% flu shot in 2002–2003	54.43	63.08	41.98
Months with dual eligibility (24 max)	5.8 (10.1)	4.1 (9.0)	8.2 (11.2)
Months with health maintenance organization coverage 2000–2001	1.5 (5.3)	1.5 (5.3)	1.5 (5.2)
% personal disability: Yes	6.18	5.52	7.13
Distance to closest mammography facility	1.9 (3.9)	1.9 (3.9)	2.0 (4.1)
% changed ZIP code of address between 2002 and 2003	3.63	2.46	5.33
% with other types of cancer	2.73	2.54	2.99
Race or ethnicity			
% Asian	6.88	5.03	9.56
% African American	4.88	3.85	6.35
% Hispanic	4.09	3.22	5.35
% Native Americans/Alaskan Natives	0.32	0.27	0.40
% Other	3.10	2.71	3.67
% White	80.73	84.92	74.67

Table 4
Sample Statistics for Area-Level Contextual Variables, by Area Type

Variable	ZCTA (<i>n</i> = 1,450)		PCSA (<i>n</i> = 333)		MSSA(<i>n</i> = 515)		County (<i>n</i> = 57)	
	M	SD	M	SD	M	SD	M	SD
Isolation index for African Americans	0.045	0.087	0.049	0.078	0.076	0.109	0.087	0.091
Isolation index for Hispanics	0.252	0.225	0.312	0.237	0.332	0.231	0.317	0.193
Isolation index for Asians	0.076	0.106	0.067	0.085	0.101	0.118	0.094	0.102
Isolation index for American Indians and Alaska Natives	0.012	0.034	0.022	0.057	0.018	0.042	0.035	0.055
Isolation index for Pacific Islanders	0.002	0.004	0.003	0.004	0.007	0.009	0.009	0.010
Proportion of workforce commuting more than 60 min each way to work	0.108	0.071	0.102	0.059	0.103	0.059	0.094	0.046
Proportion of elderly persons with little or no English language ability	0.287	0.208	0.287	0.168	0.318	0.171	0.265	0.141
Proportion of elderly persons in poverty	0.079	0.057	0.079	0.038	0.082	0.041	0.070	0.021
Proportion of female elderly persons living alone	0.309	0.131	0.318	0.067	0.309	0.073	0.334	0.035
Proportion of population immigrating into the United States between 1995 and 2000	0.165	0.106	0.179	0.069	0.176	0.066	0.185	0.042
No. mammography facilities per 1,000 elderly women	0.317	0.694	1.383	3.216	1.408	7.657	1.559	1.890
No. oncologists per 1,000 elderly persons	0.121	0.061	0.105	0.060	0.117	0.062	0.084	0.074
No. nurses per 1,000 elderly persons	69,163	93,611	54,726	82,874	67,021	90,990	30,773	69,926
Ratio of international to U.S. medical graduates (general and specialists)	0.218	0.226	0.235	0.203	0.259	0.172	0.191	0.121
Primary care physician shortage area	0.891	0.312	0.925	0.264	0.894	0.308	0.842	0.368
Health maintenance organization penetration	0.404	0.211	0.358	0.223	0.396	0.215	0.244	0.237
Violent crime rate	3,939	1,108	4,054	1,154	4,005	1,083	4,076	1,308
Index of land use mix	0.318	0.228	0.266	0.221	0.309	0.230	0.159	0.180

Note: ZCTA = ZIP code tabulation area; PCSA = primary care service area; MSSA = medical service study area.

Table 5
Binary Probit Model (Generalized Estimating Equations Cluster Estimator): Marginal Probability Estimates and *p* Values

Variable	ZCTA Model	<i>p</i>	PCSA Model	<i>p</i>	MSSA Model	<i>p</i>	County Model	<i>p</i>
Individual characteristics								
Enabling								
Disability as the original reason for Medicare entitlement	-0.046	0.00	-0.046	0.00	-0.046	0.00	-0.047	0.00
Moved to new residence in 2002–2003	-0.061	0.00	-0.061	0.00	-0.061	0.00	-0.059	0.00
Months enrolled in Medicare HMO anytime 2000–2001	-0.001	0.00	-0.001	0.00	-0.001	0.00	-0.001	0.00
Months dually eligible for Medicare and Medicaid	-0.005	0.00	-0.005	0.00	-0.005	0.00	-0.006	0.00
Distance to mammography facility closest to residence	-0.001	0.04	-0.001	0.02	-0.001	0.06	-0.001	0.00
Predisposing								
Age in 2002	-0.015	0.00	-0.014	0.00	-0.014	0.00	-0.014	0.00
Asian	-0.066	0.00	-0.068	0.00	-0.066	0.00	-0.076	0.00
African American	-0.018	0.05	-0.030	0.00	-0.022	0.01	-0.049	0.00
Hispanic	-0.003	0.70	-0.003	0.77	-0.000	0.99	-0.016	0.07
American Indian/Alaska Native	-0.068	0.02	-0.066	0.02	-0.059	0.04	-0.066	0.02
Other race or ethnicity	-0.065	0.00	-0.068	0.00	-0.066	0.00	-0.075	0.00
Need								
Diagnosed with other cancer	0.067	0.00	0.068	0.00	0.068	0.00	0.068	0.00
Diagnosed with breast cancer	0.299	0.00	0.300	0.00	0.299	0.00	0.301	0.00
Obtained flu shot from doctor	0.171	0.00	0.171	0.00	0.171	0.00	0.173	0.00
Interpersonal/proximate factors (ZCTA, PCSA, MSSA, County)								
Social integration and support								
Isolation index for African Americans	-0.057	0.01	-0.041	0.07	-0.039	0.02	-0.032	0.34
Isolation index for Hispanics	-0.039	0.00	-0.036	0.02	-0.029	0.03	-0.037	0.09
Isolation index for Asians	0.004	0.85	-0.002	0.94	-0.015	0.43	0.027	0.60
Isolation index for American Indians and Alaska Natives	0.276	0.04	0.156	0.03	0.018	0.86	0.182	0.00
Isolation index for Pacific Islanders	-0.705	0.13	-0.105	0.77	-0.079	0.68	0.406	0.25
Stressor: Proportion commuting 60 min or more each way to work	-0.125	0.00	-0.145	0.00	-0.157	0.00	-0.143	0.04
Social/cultural cohesion								

Variable	ZCTA Model	p	PCSA Model	p	MSSA Model	p	County Model	p
Proportion of elderly persons with little or no English language ability	-0.030	0.06	-0.032	0.22	-0.018	0.41	-0.070	0.16
Proportion of population immigrating into the United States between 1995 and 2000	-0.016	0.56	0.055	0.23	-0.010	0.81	0.058	0.56
Stressor, living conditions								
Proportion of female elderly persons living alone	0.063	0.01	0.102	0.02	0.085	0.01	0.100	0.45
Proportion of elderly persons in poverty	-0.304	0.00	-0.317	0.00	-0.377	0.00	-0.150	0.64
Intermediate/community factors (County)								
Health care system								
No. mammography facilities per 1,000 elderly women	0.001	0.66	0.000	0.82	0.001	0.67	0.086	0.08
No. oncologists per 1,000 elderly persons	-0.033	0.34	0.002	0.96	-0.025	0.49	-0.036	0.45
No. nurses per 10,000 elderly persons	0.001	0.01	0.001	0.01	0.001	0.03	0.001	0.04
Ratio of international medical graduates to U.S. medical graduates	0.036	0.15	0.064	0.02	0.041	0.12	0.000	0.96
Primary care physician shortage area	0.003	0.62	0.007	0.31	0.006	0.37	0.007	0.40
Health maintenance organization penetration	0.022	0.12	0.020	0.17	0.019	0.19	0.017	0.32
Social context: Violent crime rate	-0.009	0.00	-0.009	0.00	-0.008	0.00	-0.009	0.01
Physical environment: Index of land use mix	-0.064	0.00	-0.065	0.00	-0.059	0.00	-0.046	0.03

Note: $n = 70,129$ women. Prediction success rate, 71%; specificity (correct, no use), 60%; sensitivity (correct, yes use), 80%. ZCTA = ZIP code tabulation area; PCSA = primary care service areas; MSSA = medical service study areas.