Craig Duncan
and Kelvyn Jones

Using Multilevel Models to Model Heterogeneity: Potential and Pitfalls

Within the last few years, geographers and researchers in other cognate disciplines with geographic concerns have begun to use multilevel models. While there are several useful existing introductory accounts of these models in the geographical literature, this paper seeks to extend them in three main ways to clarify and emphasize further the substantial opportunities they afford. First, it focuses on how multilevel models are centrally concerned with modeling population heterogeneity as a function of predictor variables. Second, it considers and illustrates a number of specific interpretive issues that can arise when conducting multilevel analyses of place effects. Lastly, it traces some more general, conceptual issues surrounding the use of multilevel models in geographical research. The arguments made are illustrated through an analysis of variations in drinking behavior using data from a typically complex, large-scale survey; particular attention is given to the inclusion of categorical predictors.

Multilevel models are now being used by geographers in certain subdisciplinary areas, most especially political (for example, Jones, Johnston, and Pattie 1992; Carmines, Huckfeldt, and McCurley 1995; Charnock 1996) and health geography (for example, Congdon 1995; Duncan, Jones, and Moon 1995; Gould and Jones 1996; Langford and Bentham 1997; Shouls, Congdon, and Curtis 1996; Verheij, de Bakker, and Groenewegen 1999). Such models have also been taken up by researchers in other cognate disciplines with intersecting geographic interests and concerns. In health studies, for example, they have been employed by epidemiologists, public health practitioners, community psychologists/health promotion specialists and health economists (for example, Von Korff et al. 1992; Hedeker et al. 1994; Carr-Hill et al. 1994; DiezRoux 1998; Reijneveld 1998).

The key motivating reason for using multilevel models in this work has been to investigate the extent and nature of spatial variations in individual outcome measures. More specifically, while often there is unequivocal evidence that individual outcomes are different in different places, the source of such differences remains far from clear.

The authors thank the ESRC Data Archive at the University of Essex for supplying the *Health and Lifestyle Survey* Data (Cox 1988).

Craig Duncan is a research fellow and Kelvyn Jones is a professor in the Institute for the Geography of Health, University of Portsmouth. E-mail: craig.duncan@port.ac.uk. kelvyn.jones@port.ac.uk

Geographical Analysis, Vol. 32, No. 4 (October 2000) The Ohio State University
Submitted: 10/21/98. Revised version accepted: 3/31/00.

In Britain, for example, it is well established that there is a North-South divide in life expectancy with those in the North tending to live shorter lives (Britton 1990). One explanation for such a difference might be "compositional" effects—there is a clustering of people in the North who, because of individual characteristics (for example, low social class), are more likely to die younger. On the other hand, they may arise from "contextual," "area,"or "ecological" effects associated with the characteristics of places themselves rather than just the type(s) of people that they contain.

One obvious research question is the relative contribution of compositional and contextual factors. Due to several technical reasons, this is better answered using multilevel models rather than traditional single-level techniques (DiPrete and Forristal 1994; Hox and Kreft 1994). Multilevel modeling is, however, more than simply a means of separating compositional and contextual effects in overall terms. First, it provides a way of assessing for which types of people contextual effects matter. Instead of there being simple, overall differences between places, there may be important people-place interactions: contextual effects may not be the same for all types of people. Second, it offers a way of showing whether the differences between individuals themselves are too complex to reduce to simple summary "averages." Multilevel models are, therefore, a means of investigating complex between-place and between-people differences. They provide a way of explicitly modeling heterogeneity.

This paper seeks to clarify and elaborate this view and consider the potential and opportunity that multilevel modeling holds. It is not an introductory account, but a development of such accounts that already exist in the geographical literature (for example, Jones 1991; Jones and Duncan 1996; Bullen, Jones, and Duncan 1997). This development occurs in three main ways and corresponds with how the rest of the paper is structured. First, the paper focuses explicitly on how multilevel models are centrally concerned with modeling population heterogeneity as a function of predictor variables. Second, it outlines a number of specific interpretive issues that can arise when conducting multilevel analyses of place effects. Each is illustrated through an analysis of variations in drinking behavior using data from a typically complex, large-scale data set and one of the most popular specialized software packages. Third, it traces some important general conceptual issues relating to the use of multilevel models in geographical research. This emphasizes further the capacity and usefulness of the technique as well as signaling broader points of debate and reflection.

MULTILEVEL ANALYSIS: MODELING AVERAGES, MODELING HETEROGENEITY

We will begin by considering a simple analysis of drinking behavior. More specifically, motivated by recent arguments for a contextual understanding of health and health behavior (Blaxter 1990), our concern is with establishing how alcohol consumption differs between people of different ages and whether geographies of drinking relate to people's age. We begin with a single-level regression model:

$$y_i = \beta_0 x_{0i} + \beta_1 x_{1i} + (\varepsilon_i) \tag{1}$$

where y_i is the number of units of alcohol consumed in a week by an individual, i, x_0 is a set of 1s to represent the constant, and x_1 is the continuous predictor variable, age. In such a model, if the predictor variable age, x_1, is centered about its mean, β_0, the intercept, gives the average weekly alcohol consumption for a person of average age. The slope, β_1, gives the average change in alcohol consumption for a unit change in age.

This model provides estimates of the "average" age-alcohol relation, but it does not allow for heterogeneity either between places nor between individuals within places. As the two parameters, β_0 and β_1, are fixed—that is, they are not subscripted and only take one value—the model assumes there is one "universal" alcohol consumption-age relationship: thus, differences between places in the alcohol-age relation are not permitted. The model also does not allow heterogeneity between individuals. While the parameter, ε_i, captures the differences between individuals not accounted for by the fixed parameter, age, it is summarized by a single variance term, σ^2_ε. Such homoskedastic assumptions may be quite unrealisitic. People of different ages may be differentially variable in their alcohol consumption. While young people may have similar drinking habits, older people may be much more variable; put another way, between-individual variation may change according to age.

In contrast to the single-level model outlined above, multilevel models are concerned with modeling both the average and the variation around the average. To do this, they consist of two sets of parameters: those summarizing the overall, average relationship(s); those summarizing the variation around the average at both the level of places and people. This latter set come to form equations that give the total amount of variation at a level as a function of predictor variables. Returning to our example, we could specify the variation in alcohol consumption between places, j, as a quadratic function of age, x_1, by writing the equation:

$$\textit{variation between places} = \gamma_0 x_{0ij} + \gamma_1 x_{1ij} + \gamma_2 x^2_{1ij} \,. \tag{2}$$

We could also specify the variation between individuals as a quadratic function of age, x_1, by writing the equation:

$$\textit{variation between individuals} = \lambda_0 x_{0ij} + \lambda_1 x_{1ij} + \lambda_2 x^2_{1ij} \,. \tag{3}$$

Each equation consists of three parameters, one of which is associated with the constant, x_0, while the remaining two are associated with age, x_1. Such equations are consistent with the following fully specified multilevel model:

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (\mu_{0j} x_{0ij} + \mu_{1j} x_{1ij} + \varepsilon_{0ij} x_{0ij} + \varepsilon_{1ij} x_{1ij}) \,. \tag{4}$$

In this model, where i represents individuals and j places, the parameters β_0 and β_1 are fixed and give the average alcohol consumption-age relationship. The remaining subscripted terms in the brackets are random (where random means allowed to vary) and represent the differences in alcohol consumption between places and between individuals. In terms of the former, shown by the parameters with a single j subscript for places, there are two terms—μ_{0j} associated with x_{0ij}, and μ_{1j} associated with x_{1ij}.[1] Making the usual IID assumptions, these terms can be summarized in a set of variance-covariance terms: $\sigma^2_{\mu 0}$, $\sigma^2_{\mu 1}$, and $\sigma_{\mu 0 \mu 1}$. Following a well-known result (Weisberg 1980), the combined variability for two random variables is given by:

$$\mathrm{var}(\mu_{0j}, \mu_{1j}) = \sigma^2_{\mu 0} x_{0ij} + 2\sigma_{\mu 0 \mu 1} x_{1ij} + \sigma^2_{\mu 1} x^2_{1ij} \,. \tag{5}$$

Thus, the total between-place variation is a quadratic function of age based on the variance term for each variable and its covariance. This corresponds with equation (2) such that γ_0 represents $\sigma^2_{\mu 0}$, γ_1 is twice the covariance $\sigma_{\mu 0 \mu 1}$, and γ_2 is $\sigma^2_{\mu 1}$.

In terms of differences between individuals, there are also two terms. Thus, making the same assumptions and following a similar procedure, we obtain:

[1]In this case, both parameters have a straightforward interpretation: μ_{0j} is the difference in alcohol consumption for a person of average age in a particular place from the national value, β_0, while μ_{1j} represents the extent to which the place-specific alcohol-age relation differs from the national relation, β_1.

$$\text{var}(\varepsilon_{0ij}, \varepsilon_{1ij}) = \sigma^2_{\varepsilon 0} x_{0ij} + 2\sigma_{\varepsilon 0 \varepsilon 1} x_{1ij} + \sigma^2_{\varepsilon 1} x^2_{1ij} \,. \tag{6}$$

The total between-individual variation is a quadratic function of age and, so, corresponds with equation (3), such that λ_0 represents $\sigma^2_{\varepsilon 0}$, λ_1 is twice the covariance $\sigma_{\varepsilon 0 \varepsilon 1}$, and λ_2 is $\sigma^2_{\varepsilon 1}$.

In summary, the fixed parameters (β_0, β_1) give the overall, average alcohol consumption–age relationship while the random parameters combine to form quadratic functions representing the differences between places and between individuals.[2] Two important points follow from this. First, the size and magnitude of the random parameters reflect the particular relationships found. As an illustration of this, Figure 1 graphs one set of possible results. Figure 1(a) shows the shape of the between-place variation in relation to age; Figure 1(b) shows the shape of the between-individual variation; while Figure 1(c) plots the overall relationship (solid line) and the corresponding 95 percent confidence intervals for the place-level population heterogeneity (dashed lines): these are approximated as the fitted line based on the fixed-part terms plus and minus 1.96 times the square root of the estimated between-place variation around the line. From the graphs, we see that both between-place variation (1a) and between-individual variation (1b) are a *decelerating* function of age. Such a situation would occur when all variances are nonzero and both covariances are negative; it would imply that older people have more similar drinking habits and that between-place differences are smaller for such people. As (1c) shows, these differences occur around an overall relationship in which older people drink less (a positive fixed intercept and a negative fixed slope). This graph also confirms that in the population the greatest differences between places are for the young. The second point is that the functions estimated are conditional on all the other parameters in the model. Thus, when a series of fixed-part predictors describing individual characteristics are included (for example, gender, social class, educational background, housing tenure, and age), we obtain estimates of between-place heterogeneity that take account of compositional factors. In other words, we have estimates of contextual effects.

Specifying the variation between individuals and between places as quadratic functions may not always be appropriate. Rather than between-place or between-individual differences increasing (or decreasing) at an accelerating (or decelerating) rate, they may change in a linear fashion. To handle this situation, rather than specifying quadratic variance functions as in (2) and (3) we would specify linear ones:

$$\textit{variation between places} = \gamma x_{0ij} + \gamma x_{1ij} \,; \tag{7}$$

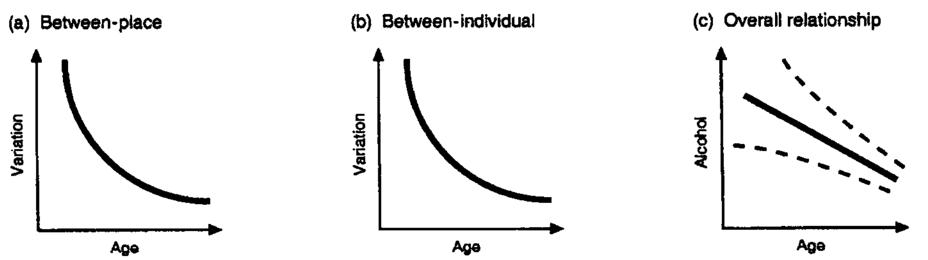

FIG. 1. Results Based on Quadratic Functions

[2]The sign of the covariance term(s) is obviously of key interpretative importance. If it is positive, variation increases as the predictor variable increases. If it is negative, variation decreases as the predictor variable increases.

$$\textit{variation between individuals} = \lambda x_{0ij} + \lambda x_{1ij}\,. \tag{8}$$

Although we would write the same overall model as equation (4) we would only estimate one variance and one covariance at each level rather than a full set of variances-covariances. Thus, we would specify:

$$\mathrm{var}(\mu_{0j}, \mu_{1j}) = \sigma^2_{\mu 0} x_{0ij} + 2\sigma_{\mu 0 \mu 1} x_{1ij}\,; \tag{9}$$

$$\mathrm{var}(\varepsilon_{0ij}, \varepsilon_{1ij}) = \sigma^2_{\varepsilon 0} x_{0ij} + 2\sigma_{\varepsilon 0 \varepsilon 1} x_{1ij} \tag{10}$$

where, as before, the parameters here correspond with those in the equivalent equations, (7) and (8).

Although estimating a covariance when there is only one variance might seem contradictory, such a specification is entirely feasible (Goldstein et al. 1998) and usefully emphasizes the "true" meaning of random parameters. While on some occasions [for example, equation (5) and equation (13) later] they are individual terms with singular, accepted meanings—variances and covariances—thinking of them more generally as constitutive elements of overall variance functions is more accurate [as given by equations (2) and (3) (quadratic) and (7) and (8) (linear)].[3] By extension, it is the functional form as a whole that has actual substantive meaning. Researchers should not simply look at individual parameters [for example, the $\sigma^2_{\mu 0}$ and $\sigma^2_{\mu 1}$ of equation (5)] but need to consider the estimates for the *function* as a whole.

This deeper understanding has been downplayed in most existing accounts of multilevel analysis in the geographical literature. In such accounts, notions of higher-level (between-place) variation have been developed through ideas of distributions of place-specific slopes and intercepts. While a useful heuristic device for introductory purposes, such ideas can lead to a misplaced emphasis on random parameters having specific individual meaning (varying slopes, varying intercepts) and relating to specific individual places (those sampled). The perspective developed here has emphasized that parameters must be seen within overall variance functions. From such a perspective, not only does the linear formulation appear quite reasonable, but the whole idea of between-individual variation also makes more sense since specific level-1 slopes and intercepts [ε_{0ij} and ε_{1ij} in equation (4)] are never meaningful.

Figure 2 illustrates one set of possible results for a model based on linear variance functions. For both between-place variation (2a) and between-individual variation (2b), these are *negative*. As before, they would be based on nonzero variances and

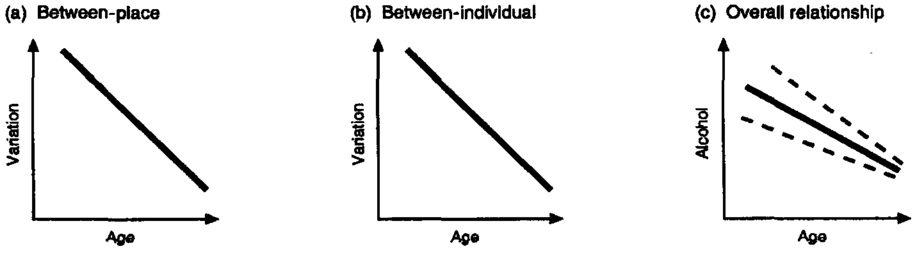

FIG. 2. Results Based on Linear Functions

[3]Within the multilevel literature, it is usual to use the terms variances and covariances throughout though quotes are put around "covariance" to show when this term is a convenience. In the remainder of this paper, we follow this convention.

negative "covariances," though now there would only be two of the former, one at each level. The differences again occur around an overall relationship in which older people drink less (2c).

Rather than specify quadratic or linear functions, it may, of course, be appropriate to specify constant functions so that the between-place and between-individual variation is unchanging with age. Thus, we would specify:

$$\textit{variation between places} = \gamma x_{0ij}\ ; \tag{11}$$

$$\textit{variation between individuals} = \lambda x_{0ij} \tag{12}$$

which would correspond with

$$\sigma^2_{\mu 0} x_{0ij} \tag{13}$$

and

$$\sigma^2_{\varepsilon 0} x_{0ij} \tag{14}$$

while the overall model would be simplified to

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (\mu_{0j} x_{0ij} + \varepsilon_{0ij} x_{0ij})\ . \tag{15}$$

In this most simple case, therefore, the variance function at each level consists of only one parameter associated with x_{0ij} since the differences are unchanging with age. Regarding equation (13), there are still place differences but these are held to be the same at all ages. Put differently, there are no differential geographies of drinking according to age: places that are high for the youngest are also high for the eldest. Equation (14), meanwhile, obviously corresponds with the usual assumption of homoskedasticity in single-level regression. As before, one set of possible results for this model can be graphed and this is done in Figure 3. These confirm that while there are both differences between places (3a) and people (3b) these do not change with age. Accordingly, the 95 percent "tramlines" for place-level population heterogeneity are parallel (3c).

Crucially, since they are based on random parameters, the variance functions relate to the broader population rather than simply the specific people or places sampled. This way of handling heterogeneity is in direct contrast to long-standing techniques such as ANOVA/ANCOVA and to more recent ones such as the expansion method (Cassetti and Jones 1992), both of which can involve expanding the number of fixed-part parameters through the inclusion of dummy or indicator-coded variables to de-

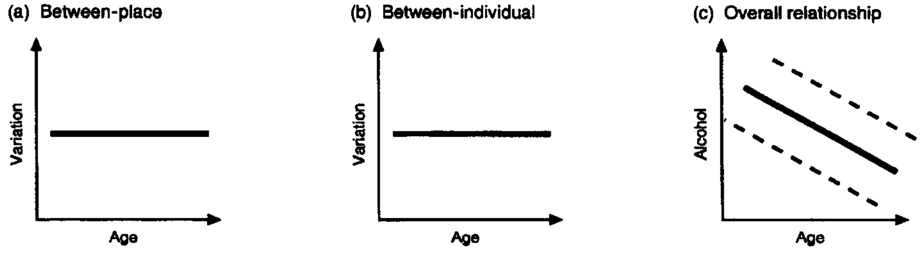

FIG. 3. Results Based on Constant Functions

note particular places. As Jones and Bullen (1994) reveal, such approaches are neither efficient nor parsimonious. More significantly, since they use traditional OLS estimation procedures, they are unable to handle between-individual heterogeneity since this violates the assumption of homoskedasticity. At the same time, inferences of between-place heterogeneity are based only on the specific places explicitly identified and not the wider population from which they are drawn. The multilevel approach, meanwhile, involves expanding the number of random terms, making iterative estimation procedures necessary. Both between-individual and between-place heterogeneity are easily accommodated and inferences are made for the broader population. It should, however, be noted that, at the place-level, *predictions* of specific relationships (the μ_{0j}s and the μ_{1j}s) can be obtained once the overall variance functions have been estimated.[4] In light of this, a multilevel estimation procedure can be viewed as a two-stage process. In the first, the overall variance functions are estimated together with the fixed-part parameters. In the second, these overall summaries are combined with place-specific information to derive predictions of place-specific intercepts and slopes. If a particular place has few observations or if there is little variation in the predictor variable(s), the predictions for such a place will be down-weighted or shrunk toward the overall fixed relationship (Morris 1983). A reliably estimated within-place relationship will, however, be largely immune to this shrinkage. In Bayesian terminology, these predictions are known as the posterior residual estimates.[5]

In summary, multilevel analysis is centrally concerned with modeling population heterogeneity, both at the level of people and at the level of places. This is achieved through the specification of variance functions based on random parameters. Nothing new or different is involved in extending them to three or more levels. Crucially, the technique contains no built-in assumptions about the heterogeneity that exists at a particular level; instead, the researcher can specify different functional forms at each level to see which receives the best empirical support from the data. While here we have shown cases where the functional form is the same at each level, in reality they might be different. Such a two-level situation is shown in Figure 4. Here, the between-place variation increases with age according to a quadratic function

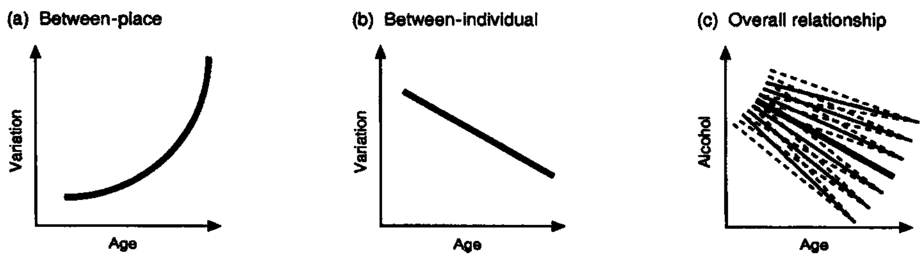

FIG. 4. Results Based on Quadratic Function at Level 2 and Linear Function at Level 1

[4]It should be noted that these have deliberately not been included on Figures 1-3 (cf. existing introductory accounts referred to earlier) so as to emphasize the key concept of modeling population heterogeneity rather than place-specific relations.

[5]By using shrinkage estimators, multilevel models have the potential to avoid the misestimation problems caused by small numbers and sampling fluctuations in traditional methods based on single-level separate regressions. It should be noted, however, that their use does not find favor with everyone and it is important to appreciate how they work; as de Leeuw and Kreft (1995, p. 184) note in relation to work on school performance, it helps to explain "the frustration of the principal of an excellent school who sees the predictions of success of her students shrunken towards the mean." A fuller discussion of shrinkage estimators within multilevel models can be found in Jones and Bullen (1994) and more attention is given to using the predictions that they generate later in this paper.

(4a) while the between-individual variation decreases with age according to a linear function (4b). In 4c, the dashed lines represent 95 percent confidence intervals for individual-level population heterogeneity around six place-specific lines and the overall line. While older people are less variable, there are greater place differences for such people.

INTERPRETING HIGHER-LEVEL HETEROGENEITY

Approaching multilevel analysis from the perspective developed in the preceding section emphasizes the way in which it offers an extremely flexible and versatile technique for investigating heterogeneity. Such flexibility and versatility does bring with it certain demands and consequences. In this section, we use a "real"[6] multilevel analysis of alcohol consumption to outline and illustrate three particular situations that researchers should be mindful of. From the outset, we should make it clear that this is not done to diminish the value of multilevel modeling for geographical research; rather, it is meant to help geographical researchers appreciate more fully the opportunities that the technique presents.

The analysis that we consider is based on data from the first British *Health and Lifestyle Survey (HALS1)* (Cox et al. 1987). This is a complex, large-scale, hierarchically structured data set that, as well as having been widely used (for example, Humphreys and Carr-Hill 1991; Duncan, Jones, and Moon 1993, Lewis and Booth 1994; Mitchell et al. 1998), is typical of survey data sets used in other research areas. It is based on a multistage (hierarchical) sampling design of individuals within electoral wards (small administrative units) within constituencies (larger political units) which we have further nested within regions. Thus, in the present analysis, three-level models are specified based on a structure consisting of 6,211 individuals at level 1 (all those who completed a drink diary), nested within 396 electoral wards at level 2, nested within 22 regions at level 3 as defined by *The Economist* classification (Johnston, Pattie, and Allsopp 1988). The response variable is continuous and represents the number of units of alcohol consumed in a week. The analysis was performed using the *MLwiN* software package (Goldstein et al. 1998).

Different Implementations

One usual starting point in multilevel analyses is to estimate what is known as a null model. This is an extremely simple model and is similar to that given in equation (15), that is, the variation at each level is summarized by a single parameter. There are, however, no predictor variables; only the constant is included. Thus, as well as estimating the grand mean, such models provide an estimate of the total variation at each level without taking account of population composition. As already stated, if a range of individual predictor variables is then included, it becomes possible to see how compositional factors affect the degree of variation at each level. More specifically, it shows whether geographies remain that do not relate to the types of people in particular places.

Table 1 gives the results for alcohol consumption in *HALS1* before and after including a set of predictor variables for age, gender, social class, employment status, housing status, marital status, ethnicity, household income, and age leaving school. This list of variables is in accordance with those normally used in public health/epidemiological research. In contrast to econometric applications, price is not included.

[6]By "real" we mean that we have not manipulated the data in any way to produce the situations that we consider. Instead, they occurred naturally during a period of data analysis conducted by the authors. That said, however, our intention here is not to offer a definitive analysis of drinking behavior but to represent situations that may arise in any typical piece of large-scale survey analysis using multilevel techniques.

TABLE 1
Results for the Constant Variation Models without (Column A) and with (Column B) Level-1 Predictors

	(A)		(B)	
Fixed Effects				
Level-1 (Individual)				
Constant	11.88	(0.37)	5.31	(0.57)
Age			–0.10	(0.02)
Gender				
Male			11.94	(0.38)
Social Class				
I&II			–0.41	(0.52)
III nonmanual			–1.46	(0.61)
IV&V			–0.08	(0.53)
Missing			–2.42	(1.39)
Employment Status				
Unemployed			2.55	(0.87)
Housing Status				
Local Authority renter			1.63	(0.50)
Other renter			0.87	(0.72)
Missing			–4.72	(3.45)
Marital Status				
Single			2.86	(0.58)
Widowed			1.63	(0.88)
Divorced/Separated			3.89	(0.80)
Ethnicity				
Nonwhite			–7.10	(1.40)
Missing			–5.28	(2.18)
Household Income				
Low (≤£415 per month)			–1.74	(0.53)
High (≥£996 per month)			2.47	(0.59)
Missing			–0.12	(0.54)
Age Leaving School				
16			–1.21	(0.50)
Post-16			–0.85	(0.57)
Missing			–7.89	(4.54)
Random Effects				
Level 3 (Region)				
Constant, $\sigma^2_{\nu 0}$	1.83	(0.90)	1.57	(0.76)
Level 2 (Ward)				
Constant, $\sigma^2_{\mu 0}$	2.54	(1.38)	1.67	(1.11)
Level-1 (Individual)				
Constant, $\sigma^2_{\varepsilon 0}$	253.2	(4.69)	209.3	(3.88)

NOTE: Estimates represent units of alcohol; figures in parentheses are standard errors.

This may seem to be a significant omission in terms of compositional effects. From a multilevel perspective, however, the price of a good can be thought of more as an area-level characteristic rather than an individual one—places might matter not only through shaping consumption tastes/preferences but also through influencing price-setting mechanisms. While there may be accompanying endogeneity problems (an issue considered later in relation to other research), an interpretation of price as, intrinsically, local and relative does seem reasonable for our purpose here and fits with existing research on the patterning of health-related behaviors rather than the analysis of consumption functions for particular goods.

Apart from age, which was represented as a continuous variable centered about its mean, the other variables are represented by dummy/indicator-coded variables contrasted with the base category of a forty-six-year-old employed woman who left school before the age of sixteen, is married, does not belong to an ethnic minority, and lives in an owner-occupied household with average income (£416–£995 per month), the head of which is in social class III-manual. This base category represents the individual characteristics that occur most frequently among all respondents to the survey and the estimates shown in Table 1, column B, represent differences from these. To test their significance, the ratio of an estimate to its standard error can be calculated; if the result is more than ± 2, the estimate can be judged to be significantly different from zero at the 0.05 level. The findings are substantially in accordance with established research: alcohol consumption is greater for men, for the unemployed, for local authority renters, for those not married, and for those in high-income households.

Of most significance, given our focus here, however, is that including these predictors leads to a reduction in the size of both higher-level random terms. Neither higher-level term is large in relation to its standard error or to the estimate of level-1 variation and, on the basis of a χ^2 test, only the region level intercept has marginal significance ($p=0.04$).[7] According to these results, therefore, it seems that places—be they wards or regions—do not play any substantial independent role in shaping individual drinking behavior.[8] In short, there are no strong geographies of drinking arising from place-contingent social and economic processes.

Following the early applications in educational research (for example, Raudenbush and Willms 1991), it was considered that, like here, adding individual-level predictors could only lead to a reduction in higher-level variation. It is, however, possible for higher-level variation to increase after considering population composition. Thus, genuine contextual effects can be hidden or masked by not allowing for social and demographic composition just as easily as false contextual effects can be exaggerated. While this has been found in other health geographic research (Shouls, Congdon, and Curtis 1996), it is perhaps best illustrated through house prices (Jones and Bullen 1993). Places with genuinely high house prices may, for example, be characterized by smaller, and usually cheaper, properties. Thus, Kensington, and other areas like it in London are obviously expensive, but when account is taken of what is being sold there (small apartments), they are even more expensive. If there are several such areas, between-place variation can increase once composition is taken into account.

The practice of first fitting null models and then adding a range of theoretically relevant individual-level variables is, therefore, a good one as any changes in the estimates of between-place variation will reveal the precise effect of compositional factors, and, by extension, how places matter. While many researchers have, in fact, followed this procedure, it is crucial that the importance of doing so is appreciated and understood. In broader terms, what is really at issue is whether the fixed-part of the model is well specified (that is, many appropriate and relevant individual-level predictors have been included). Not only does this secure a more faithful estimate of contextual differences, it also removes support for the argument that apparent contextual effects are simply a result of misspecified individual effects (Hauser 1970).

While model B in Table 1 took account of population composition, it remained extremely simple in terms of modeling between-place heterogeneity: that is, only a constant function was specified. Thus, according to this model the variation between places is the same for all types of people. Such an assumption should not remain untested, for it can mean that important place effects remain hidden. Unfortunately, many applications, particularly, but not exclusively, the earlier ones, have not taken advantage of this important, additional capacity. Earlier, we considered only the case

[7]For parameters in the random part of the model, the simple significance test just outlined should only be regarded as a rough rule of thumb. A more reliable testing procedure based on χ^2 statistics is available within *MLwiN* and has been used here.

[8]When the response is continuous, it is possible to calculate the proportion of total variation attributed to each level by dividing the amount at that level by the total amount across all levels; this is also a measure of the degree of intraclass autocorrelation. Contrary to Duncan, Jones, and Moon (1993), this procedure cannot be followed when multilevel logit models are fitted to binomial responses since the level-1 variation is calibrated on a different scale (proportional) to the higher-level variation (logit). Snijders and Bosker (1999) do, however, provide alternative methods of calculation.

where complex between-place heterogeneity is modeled as a function of a continuous individual predictor, age. Now, we consider categorical variables. As an example, we relate drinking behavior to social class where people are classified according to a simple upper- and lower-social class dichotomy.[9] Using dummy/indicator-coding such that x_0 is defined as a set of 1s as usual, while x_1 is set to 1 when an individual belongs to the high-social-class category and is 0 otherwise, we can follow the same procedure as before to produce a model in which we allow for complex between-place heterogeneity on the basis of social class, such that

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (\mu_{0j} x_{0ij} + \mu_{1j} x_{1ij} + \varepsilon_{ij}) \,. \qquad (16)^{10}$$

Again, the fixed parameters give the average values. Given the coding procedure used, β_0 will represent the average alcohol consumption for a low social class individual while β_1 will give the average *differential* in alcohol consumption for high social class individuals. As before, two random variables summarize the between-place variation and we can specify either a full quadratic or a linear function. In the case of the former, we would estimate three parameters such that, given the coding procedure used, the variation for low-social-class people would be estimated directly by $\sigma^2_{\mu 0}$ while the degree of variation for high social class people would be given by the full function:[11]

$$\sigma^2_{\mu 0} x_{0ij} + 2\sigma_{\mu 0 \mu 1} x_{1ij} + \sigma^2_{\mu 1} x^2_{1ij} \qquad (17)$$

In the case of the linear model, we would estimate only two parameters: variation for low-social-class people would again be given directly by $\sigma^2_{\mu 0}$, while variation for high-social-class people would now be given by the function

$$\sigma^2_{\mu 0} x_{0ij} + 2\sigma_{\mu 0 \mu 1} x_{1ij} \,. \qquad (18)$$

Of course, a constant function could also be specified by estimating only a single term, $\sigma^2_{\mu 0}$, which would take us back to a model like equation (15) and the presumption that between-place variation is unchanging by social class type. Due to the coding procedure used, parameters for the category representing the constant (low social class) are estimated directly, while those for the other category (high social class) are estimated as differentials from these. Thus, the average alcohol consumption for high social class is not simply β_1 but $\beta_0 + \beta_1$. At the same time, the between-place variation for high social class is not simply $\sigma^2_{\mu 1}$ but the appropriate linear or quadratic function as a whole.

Returning to the analysis of alcohol consumption in *HALS1*, whilst between-place variation could be considered in relation to any of the individual level predictors we will explore differences on the basis of social class categorizations. Consequently, a series of models are formed based on the procedure just outlined for specifying be-

[9]It is also possible to consider place effects in terms of subgroups of the population based on classifications consisting of more than two categories (for example, low/middle/high) or more than one characteristic (for example, social class and age). For reasons of clarity and space, we outline only the simpler cases in this paper.

[10]To mirror the empirical analysis, this model assumes there is no complex heterogeneity between people; this assumption is relaxed later.

[11]As with continuous predictors, the covariance term is of key interpretive importance. If it is positive, then high-social-class people will be more variable than low-social-class, but places that have high alcohol consumption for one type will tend to have high consumption for the other. A negative covariance could imply that either high-social-class people are less variable in their alcohol consumption across places or that places that have high consumption for one type have relatively low consumption for the other. The exact interpretation can only be reached by considering all three random terms simultaneously.

tween-place heterogeneity as complex, nonconstant functions of dummy/indicator-coded categorical predictor variables. More specifically, linear and quadratic functions based on a series of dummy/indicator variables showing four categories of social class—I&II, III manual, IV&V, and Missing—are specified while the constant represents the base category, III manual.[12]

After exploring a series of different models, it is apparent that at the ward level the most parsimonious specification involves fitting a quadratic function for social class IV&V, and a more simple linear one for the other dummy/indicator-coded categories. The full results for this model are given in Table 2. As can be seen, all of the level-2

TABLE 2
Results for the Complex Level-2 Variation Model

Fixed Effects		
Level-1 (Individual)		
Constant	5.31	(0.57)
Age	–0.09	(0.01)
Gender		
Male	11.93	(0.38)
Social Class		
I&II	–0.54	(0.51)
III nonmanual	–1.48	(0.58)
IV&V	–0.16	(0.58)
Missing	–2.48	(1.36)
Employment Status		
Unemployed	2.45	(0.87)
Housing Status		
Local Authority renter	1.45	(0.51)
Other renter	1.12	(0.71)
Missing	–5.15	(3.41)
Marital Status		
Single	2.87	(0.57)
Widowed	1.60	(0.88)
Divorced/Separated	3.98	(0.80)
Ethnicity		
Nonwhite	–6.95	(1.39)
Missing	–5.22	(2.17)
Household Income		
Low (≤ £415 per month)	–1.66	(0.53)
High (≥ £996 per month)	2.46	(0.58)
Missing	–0.06	(0.53)
Age Leaving School		
16	–1.20	(0.49)
Post-16	–0.63	(0.56)
Missing	–8.08	(4.44)
Random Effects		
Level 3 (Region)		
Constant, $\sigma^2_{\theta 0}$	1.20	(0.61)
Level-2 Random Part (Ward)		
Constant, $\sigma^2_{\mu 0}$	6.30	(2.32)
Constant/I&II, $\sigma_{\mu 0 \mu 1}$	–4.46	(1.55)
Constant/III nonmanual, $\sigma_{\mu 0 \mu 2}$	–8.83	(1.82)
Constant/IV&V, $\sigma_{\mu 0 \mu 3}$	0.27	(2.38)
IV/V, $\sigma^2_{\mu 3}$	16.33	(6.95)
Constant/Missing, $\sigma_{\mu 0 \mu 4}$	–4.50	(4.26)
Level-1 Random Part (Individual)		
Constant, $\sigma^2_{\varepsilon 0}$	206.8	(3.92)

NOTE: Estimates represent units of alcohol; figures in parentheses are standard errors.

[12]The social class categorization used was based on that applied in the survey that follows the Registrar General's classification. Social class III manual was chosen as the base category as it was the most prevalent group amongst the respondents. The small number of individuals (2.1 percent) who were in full-time education or the military, were unclassifiable, or had never been in an occupation formed a separate group which is here given the title Missing.

random terms except one display a range of values substantially different from zero. Furthermore, on the basis of a χ^2 test, all of the terms are significant except the two "covariances" between the base category, III manual, and IV&V and Missing (the terms $\sigma_{\mu0\mu3}$ and $\sigma_{\mu0\mu4}$ in Table 2). Thus, while on average social class differences are not marked (see fixed effects), there are important differences between places. Consequently, if we had fitted only a single-level model we would have found a set of average relationships suggesting class is unrelated to alcohol consumption, but these would hide important place-specific differences.

Places (wards) may, then, play an independent role in shaping alcohol consumption: certain social class groups are not drinking the same amount everywhere. These results will be interpreted further in the next section. For the moment, however, let us emphasize the broader issue articulated in this section. Different implementations can generate quite different interpretations. If only simple models based on constant functions are estimated, the results obtained can be misleading. On their own, simple models do not show the need for more complex ones. More complex models, depending on the results obtained, however, will show the value of more simple ones. In light of this, therefore, researchers need to ensure that their models are realistically complex (Best et al. 1996) without, at the same time, being overly complex (Jones and Duncan 1996, p. 103).

Different Software Packages

As already noted, when using categorical predictors based on dummy/indicator coding, values for the base category are estimated directly while those for the remaining categories are estimated as differentials from these. Using x_0 to represent the base category, III manual, and x_1 to x_4 to represent the remaining categories—I&II, III nonmanual, IV&V, and Missing, respectively—the between-ward variation for each social class category according to the model in Table 2 is given by

$$\text{III manual:} \quad \sigma^2_{\mu0}x_{0ij}\ ; \tag{19}$$

$$\text{I\&II:} \quad \sigma^2_{\mu0}x_{0ij} + 2\sigma_{\mu0\mu1}x_{1ij}\ (\text{linear})\ ; \tag{20}$$

$$\text{III nonmanual:} \quad \sigma^2_{\mu0}x_{0ij} + 2\sigma_{\mu0\mu2}x_{2ij}\ (\text{linear})\ ; \tag{21}$$

$$\text{IV\&V:} \quad \sigma^2_{\mu0}x_{0ij} + 2\sigma_{\mu0\mu3}x_{3ij} + \sigma^2_{\mu3}x_{1ij}\ (\text{quadratic})\ ; \tag{22}$$

$$\text{Missing:} \quad \sigma^2_{\mu0}x_{0ij} + 2\sigma_{\mu0\mu4}x_{4ij}\ (\text{linear})\ . \tag{23}$$

Given that the predictor variables only take the value 0 or 1, the equations above reduce to the appropriate random parameter(s). Thus, for the model in Table 2 we obtain the results shown in Table 3.

TABLE 3
Between-Ward Variability in Alcohol Consumption for the Different Social-Class Categories

Social Class Category	Total Variation	Relevant Terms (Functional Form)
III manual	6.30	$\sigma^2_{\mu0}$
I&II	–2.62	$\sigma^2_{\mu0} + 2\sigma_{\mu0\mu1}$ (Linear)
III nonmanual	–11.36	$\sigma^2_{\mu0} + 2\sigma_{\mu0\mu2}$ (Linear)
IV&V	23.17	$\sigma^2_{\mu0} + 2\sigma_{\mu0\mu3} + \sigma^2_{\mu3}$ (Quadratic)
Missing	–2.70	$\sigma^2_{\mu0} + 2\sigma_{\mu0\mu4}$ (Linear)

These results confirm the interpretation just made. Certain groups—social class IV&V and III manual—seem to display substantial degrees of between-ward variability in alcohol consumption. Unfortunately, however, such results cannot be accepted readily since the total variation for three social class categories is estimated as negative. In substantive terms, negative values like these do not have any sensible meaning; instead, they are a sign of model misspecification. Such results also draw attention to some important differences between two of the main specialized software packages currently available.

In the present case, the results arise because of large negative values for the parameters we have been calling "covariances." It is also possible that the parameters we have been calling variances might (also) be negative. Sometimes, either situation might not indicate a problem because, as we have kept emphasizing, it is the function as a whole that matters. Thus, while specific parts of the function, including the variances, may take negative values, the overall total variation could remain positive. Crucially, however, different software packages have different ways of handling individual negative variance parameters. In *HLM* (Bryk et al. 1986), one of the two main packages and the one widely used in North America, they are simply not allowed. Although understandable given that such results might appear nonsensical,[13] this does mean that researchers lose a useful diagnostic: as de Leeuw and Kreft (1995, p. 185) put it, a "better indication that something is wrong." *MLwiN*, the other main package and the one used here, meanwhile, does allow them but, in default mode, will reset them (and any associated covariances) to zero. While the user can choose to override this, no formal notification is given if this is not done and resetting is initiated automatically by the software.

Thus, although not made particularly clear in the literature to date, the issue of negative variances is an important one. *MLwiN* users, in particular, need to be aware that whenever zero estimates for higher-level variances and covariances are obtained, models should be reestimated with the "allow negative variances" option selected. Doing this will help establish whether the terms really are zero or whether they have been set to zero by the software. This situation illustrates a more general, fundamental point. Choosing which software package to use is an important decision that can have a large influence on the results obtained. To ensure accurate inferences are made, researchers need to be aware of the way in which the particular package chosen operates. As de Leeuw and Kreft put it:

> some may think, it is (equally) irrelevant which computer program is used to compute the estimates. But this is true in the same sense that it is irrelevant which means of transportation you use to get to work. Eventually you will get there all right, no matter what means of transportation you use, but walking takes hours, the bus is unpleasant, and an old car breaks down all the time. (de Leeuw and Kreft 1995, p. 184)

Actually, there is even more to it than this. While a range of multilevel modeling software packages are presently available, only *MLwiN* and *BUGS* (Gilks et al. 1993) can perform all of the analyses discussed in this paper. Continuing de Leuuw and Kreft's metaphor, therefore, certain modes of transport will not take you to all places. A detailed overview and comparison of some of the packages available is presented by Kreft, de Leeuw, and Van der Leeden (1994) and a full discussion is also available in de Leeuw and Kreft (1995) though these are now becoming dated given the new functionality of more recent software releases.

Different Overall Specifications

Negative variances/variance functions are, therefore, an important indicator that population heterogeneity is not being correctly specified. One possible cause is that

[13]From the perspective developed here, such confusion should be much less likely.

no account is taken of complex forms of variation between individuals. Ignoring such variation can also lead to exaggerated estimates of place differences as heterogeneity between different levels can be confounded: what may seem between-place variation may be within-place, between-individual variation. The model's overall specification at each and every level, therefore, also has important bearing on how we interpret the importance and size of place effects.

To date, the published literature overall and the geographical literature in particular has focused on the elaborations of the higher-level random terms so as to capture between-context heterogeneity. Indeed, this is all the current version of the *HLM* package can handle. Between- individual variation has, meanwhile, tended to be modeled only according to simple, constant variance functions [equation (14)]. As a matter of routine practice, however, researchers should estimate models in which between-individual heterogeneity is specified according to complex linear and quadratic functions for those individual predictor variables for which it might be expected to be substantial. If this is not done, then models are only taking into account compositional effects in terms of simple overall averages (the fixed parameters in the model) rather than also in terms of how types of people vary among themselves. If such variation is important yet remains ignored, it is another way in which compositional factors can produce artifactual place-effects.

While we have considered only the case where complex between-individual heterogeneity is modeled as a function of a continuous individual predictor, it can also be modeled as a function of categorical predictors. This presents some issues since individuals can only be in one category on any single dimension, making it impossible to estimate all three terms required by the quadratic function (Goldstein 1995). A linear model can, however, be estimated in one of two ways and we shall now show these by extending our example based on a simple high/low social class categorization.

First, we can create two dummy/indicator variables, z_{1ij} and z_{2ij}, where $z_{1ij} = 1$ if low social class, 0 if high social class and $z_{2ij} = 1$ if high social class, 0 if low social class, and specify the following functions at level 1:

$$\textit{variation (low social class)} = \lambda_1 z_{1ij} \; ; \tag{24}$$

$$\textit{variation (high social class)} = \lambda_2 z_{2ij} \; . \tag{25}$$

Two separate variances would, therefore, be estimated, giving the overall model:

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (\mu_{0j} x_{0ij} + \mu_{1j} x_{1ij} + \varepsilon_1 z_{1ij} + \varepsilon_2 z_{2ij}) \tag{26}$$

such that the variance of each group is given directly, that is, $\sigma^2_{\varepsilon 1}$ and $\sigma^2_{\varepsilon 2}$ respectively.

Alternatively, we can maintain a dummy/indicator coding approach and model the differences between categories by specifying what is the level-1 equivalent of the linear formulation for between-place heterogeneity based on categorical predictors described earlier. Thus, we retain x_{0ij} and the dummy/indicator-coded variable, x_{1ij}, and model the total variance at level 1 as

$$\mathrm{var}(\varepsilon_{0j} x_{0ij} + \varepsilon_{1j} x_{1ij}) = \sigma^2_{\varepsilon 0} x_{0ij} + 2\sigma_{\varepsilon 0 \varepsilon 1} x_{1ij} \; . \tag{27}$$

As equation (27) shows, we omit the variance of ε_{1ij} and estimate only

$$\mathrm{var}\ (\varepsilon_{0ij}) = \sigma^2_{\varepsilon 0} \tag{28}$$

and

$$\mathrm{cov}\ (\varepsilon_{0ij}, \varepsilon_{1ij}) = \sigma_{\varepsilon 0 \varepsilon 1} \tag{29}$$

which gives the overall model:

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + (\mu_{0j} x_{0ij} + \mu_{1j} x_{1ij} + \varepsilon_0 x_{0ij} + \varepsilon_1 x_{1ij}) . \tag{30}$$

The separate between-individual variances can then be derived with the variation for low social class being given by $\sigma^2_{\varepsilon 0}$, and the variation for high social class by $\sigma^2_{\varepsilon 0} + 2\sigma_{\varepsilon 0 \varepsilon 1}$. This procedure—modeling the *difference* in variability between categories and then deriving the variances indirectly—can be easily extended to more than two categories with the variances for the additional categories being derived in the same way.[14]

The importance of considering complex forms of between-individual variability can be seen by further extending the analysis of alcohol consumption in *HALS1*. Table 4 gives the results when between-individual heterogeneity is taken into account. This is done on the basis of gender as this was found to have the most substantial effect. The second procedure summarized in equation (27) is adopted. As the results show, not only are men likely to drink more on average (the fixed differential effect for male, 11.86) they are also much more variable in their drinking habits. Using the appropriate equations, the total between-individual variation for women is 48.5, while for men it is 377.1 [48.53+(2*164.3)]. As is also evident, the results for the ward-level random-part have changed markedly. First, it can be seen that there is now only one large negative value and, for this term, zero is within its 95 percent credible interval. Thus, by taking into account between-individual variability, the model is much better specified overall and we substantially reduce the problem of negative variance functions. Second, we also see that the terms that were previously estimated as significantly positive have reduced considerably; what seemed to be between-ward variability is actually an artifact of between-individual variability. Most notably, therefore, allowing one variable (gender) to have complex variation at level 1 results in major changes in the parameters associated with another variable (social class) at level 2. In short, the technique is flexible so that cross-confounding can be considered by level and, simultaneously, by variable(s).

These results show clearly the importance of specifying models that take into account complex forms of between-individual variation. By so doing, not only do we gain a valuable insight into differences between individuals that are substantively interesting in their own right, but we are also more likely to have a well-specified model in which place effects are correctly estimated. The present analysis provides a specific illustration of the most likely outcome of a potentially general problem: the overestimation of place effects when between-individual heterogeneity is ignored.

The model's overall specification is also important in one other way. While so far we have concentrated on modeling place effects through higher-level variance functions based on random parameters, they can also be modeled by fixed parameters based on variables describing place characteristics. Importantly, there are good grounds for suggesting that such variables should always be considered. The reasons for this can be found in a paper by Macintyre, MacIver, and Sooman (1993) considering area variations in health status. As the authors emphasize, studies that attempt to classify area variations into compositional and contextual components assume that the two effects can be easily separated. Strictly speaking, however, such a separation requires that the type of areas in which people live is independent of their sociodemographic characteristics. Given the often highly differentiated sociospatial structure of civil society, this is somewhat unrealistic: certain types of people are more likely to live in certain types of area.

[14]Overall, this second method is more flexible as it is reducible and should be used for complex classifications. For further details, see Woodhouse (1995) and Bullen, Jones, and Duncan (1997).

TABLE 4
Results for the Model Including Complex Between-Individual Heterogeneity

Fixed Effects		
Level-1 (Individual)		
Constant	5.18	(0.33)
Age	–0.02	(0.01)
Gender		
Male	11.86	(0.38)
Social Class		
I&II	–0.07	(0.33)
III non-manual	–0.61	(0.39)
IV&V	–0.04	(0.36)
Missing	–1.56	(0.81)
Employment Status		
Unemployed	2.74	(0.72)
Housing Status		
Local Authority renter	0.17	(0.32)
Other renter	0.70	(0.46)
Missing	–1.93	(2.43)
Marital Status		
Single	2.27	(0.39)
Widowed	0.37	(0.52)
Divorced/Separated	2.55	(0.49)
Ethnicity		
Nonwhite	–3.15	(1.00)
Missing	–3.99	(1.44)
Household Income		
Low ($\leq$ £415 per month)	–0.89	(0.34)
High ($\geq$ £996 per month)	1.94	(0.39)
Missing	–0.12	(0.33)
Age Leaving School		
16	–0.55	(0.32)
Post-16	0.41	(0.35)
Missing	–6.94	(3.88)
Random Effects		
Level 3 (Region)		
Constant, $\sigma^2_{\theta 0}$	0.39	(0.23)
Level-2 Random Part (Ward)		
Constant, $\sigma^2_{\mu 0}$	0.26	(0.91)
Constant/I&II, $\sigma_{\mu 0 \mu 1}$	–0.20	(0.65)
Constant/III nonmanual, $\sigma_{\mu 0 \mu 2}$	1.08	(0.85)
Constant/IV&V, $\sigma_{\mu 0 \mu 3}$	0.44	(1.01)
IV&V, $\sigma^2_{\mu 3}$	1.41	(2.42)
Constant/Missing, $\sigma_{\mu 0 \mu 4}$	–2.37	(1.52)
Level-1 Random Part (Individual)		
Constant, $\sigma^2_{\varepsilon 0}$	48.53	(1.34)
Constant/Male, $\sigma_{\varepsilon 0 \varepsilon 1}$	164.3	(4.82)

NOTE: Estimates represent units of alcohol; figures in parentheses are standard errors.

This lack of independence between residential location and individual sociodemographic location has significant implications for research examining area variations in general and, by extension, the use of multilevel models in geographical research. In conceptual terms, researchers need to acknowledge that compositional and contextual effects are often likely to be fused in some sense. In practical terms, this fusion means that by including individual characteristics in multilevel models, researchers may also in some way be taking into account place characteristics. If, however, these place characteristics have not been explicitly identified, the model may overstate individual effects while place effects are understated. This situation is the equivalent of model misspecification in single level regression due to omitted variables. As Deegan (1976) and others have shown in this instance, bias occurs when omitted variables are correlated with those included. Thus, by extension here, if we omit higher-level variables related to individual characteristics, and if both are related to the response, then bias can be anticipated. More specifically, should higher-level variance terms be estimated as nonsignificant, it does not necessarily follow that there are no place effects; they may be being subsumed by the individual variables included.

Given this, researchers should routinely include relevant individual-level *and* place-level predictors. This is achieved by simply extending the fixed-part of the model to include higher-level variables.[15] Moreover, when the individual-level predictors are categorical, interaction terms can be included to see whether the effect of a higher-level variable is different for different types of people. Thus, returning to our simplified example, we could extend the last overall model [equation 30)] to include one higher-level variable, w_{ij}, measuring the percentage of low-social-class people in a place, by writing the following:

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + \alpha_1 w_{1j} x_{0ij} + \alpha_2 w_{1j} x_{1ij} + (\mu_{0j} x_{0ij} + \mu_{1j} x_{1ij} + \varepsilon_0 x_{0ij} + \varepsilon_1 x_{1ij}) \,. \tag{31}$$

In this model, α_1 would give the effect of the socioeconomic ecology of places on alcohol consumption for low-social-class people while α_2 would show how different this effect was for high-social-class people. More broadly, by explicitly recognizing the character of places in terms of their social-class profile, we are in effect recognizing that people belonging to certain social classes may live in particular types of places. As this place effect, represented by the terms α_1 and α_2, is now explicitly recognized it will no longer be confounded with the individual effects for social class, the β_0 and β_1 terms. Thus, fixed parameters associated with higher-level variables are an important part of modeling place effects. As equation (31) also shows, however, such parameters do not supplant higher-level variance functions. Instead, the latter remain and become conditional on both compositional and (fixed) contextual factors; they reflect the variation between places unaccounted for by either individual-level or place-level variables.

The possibility of misestimating place effects by omitting place-level variables from the fixed part of the model can be illustrated by returning again to the analysis of alcohol consumption in *HALS1*. As we have seen, the previous set of results (Table 4) seemed to suggest that local neighborhoods have little independent bearing on individual drinking behavior. Extending the model to include a higher-level variable, however, gives a different impression. The variable included is an index of deprivation derived from ward data from the 1981 British Census. The index can be taken to represent the broad socioeconomic ecology of the area in which respondents live; further details of how it was constructed can be found in Duncan, Jones, and Moon (1999). A full set of interaction terms is included to allow each of the social class groups to have a different ward deprivation–alcohol relationship; the results obtained when this is done are given in Table 5. The ward-alcohol relationships for different social classes are shown graphically in Figure 5. It is clear from this graph that the differences in alcohol consumption between the least (−10) and most (+10) deprived wards are quite substantial. While there is some suggestion that the trend in consumption with place-deprivation is less marked for those in social class I&II and even reverses for the miscellaneous category, the results table shows that these differentials are not significantly different from the base category of III manual. The deprivation effect, therefore, does not seem to interact with individual social class. Table 5

[15]It should be noted that including place characteristics as higher-level variables in multilevel models provides a more faithful estimate of their effect than if they are included within traditional single-level models. For full technical details see Aitken and Longford (1986, p. 15); for several empirical illustrations see Bryk and Raudenbush (1992).

TABLE 5
Results for the Model Including Higher-Level Variables

Fixed Effects		
Level-1 (Individual)		
Constant	5.18	(0.33)
Age	−0.02	(0.01)
Gender		
Male	11.86	(0.38)
Social Class		
I&II	−0.07	(0.34)
III non-manual	−0.50	(0.40)
IV&V	−0.05	(0.37)
Missing	−1.51	(0.83)
Employment Status		
Unemployed	2.72	(0.72)
Housing Status		
Local Authority renter	−0.12	(0.34)
Other renter	0.63	(0.46)
Missing	−1.87	(2.43)
Marital Status		
Single	2.26	(0.40)
Widowed	0.40	(0.52)
Divorced/Separated	2.56	(0.50)
Ethnicity		
Nonwhite	−3.20	(0.99)
Missing	−3.82	(1.44)
Household Income		
Low (≤ £415 per month)	−0.93	(0.34)
High (≥ £996 per month)	1.98	(0.39)
Missing	−0.09	(0.34)
Age Leaving School		
16	−0.52	(0.32)
Post-16	0.44	(0.35)
Missing	−7.08	(3.89)
Fixed Effects		
Level 2 (Ward)		
Deprivation	0.16	(0.07)
Deprivation*I&II	−0.13	(0.09)
Deprivation*IIInon-man	0.01	(0.10)
Deprivation*IV&V	−0.03	(0.10)
Deprivation*Missing	−0.20	0.23)
Random Effects		
Level 3 (Region)		
Constant, $\sigma^2_{\theta 0}$	0.29	(0.19)
Level-2 Random Part (Ward)		
Constant, $\sigma^2_{\mu 0}$	0.15	(0.90)
Constant/I&II, $\sigma_{\mu 0 \mu 1}$	−0.16	(0.64)
Constant/III nonmanual, $\sigma_{\mu 0 \mu 2}$	1.06	(0.84)
Constant/IV&V, $\sigma_{\mu 0 \mu 3}$	0.52	(1.00)
IV/V, $\sigma^2_{\mu 3}$	1.18	(2.40)
Constant/Missing, $\sigma_{\mu 0 \mu 4}$	−2.29	(1.50)
Level-1 Random Part (Individual)		
Constant, $\sigma^2_{\varepsilon 0}$	48.61	(1.34)
Constant/Male, $\sigma_{\varepsilon 0 \varepsilon 1}$	163.9	(4.81)

NOTE: Estimates represent units of alcohol; figures in parentheses are standard errors.

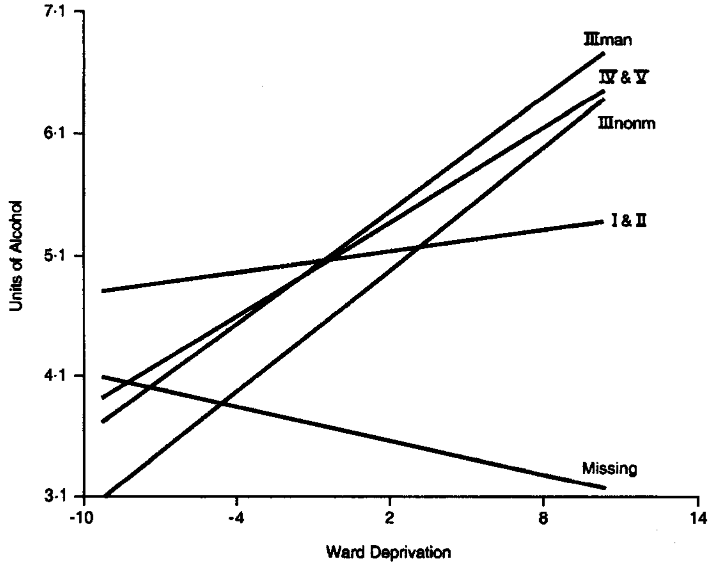

FIG. 5. Individual Social Class and Ward Deprivation Relationships for Alcohol Consumption

also confirms that individual effects can be misestimated if higher-level variables are excluded as many of these have reduced, albeit slightly.

These new results further emphasize the way in which multilevel techniques provide a flexible and powerful way of exploring complex people-place relationships. Specifically, here, they suggest that the character of the local neighborhood does play some role in shaping drinking behavior. More generally, they confirm a crucial analytical point: if higher-level variance terms are estimated as nonsignificant after including individual level variables (Table 4), it does not necessarily follow that higher-level variables should be ignored. Given this, therefore, the argument of Hauser (1970) by which place effects can always be viewed as an artifact caused by the misspecification of individual-level effects needs to be routinely applied the other way round. Omitted variable problems may apply just as much at the contextual level as at the individual one.

SOME IMPORTANT GENERAL ISSUES

While illustrated through a particular example, the specific issues considered above may arise during any multilevel analysis. Importantly, there are also other, more general, issues surrounding the use of multilevel models that require careful consideration. Until recently, researchers in many disciplines seem to have been carried away in an enthusiastic rush to use the new technique and such issues have tended to be ignored. As a series of papers in a special edition of the *Journal of Educational and Behavioral Statistics* emphasize, however, that period is now over and multilevel models have been the subject of close scrutiny and critical appraisal (see, for example, de Leeuw and Kreft 1995; Draper 1995; Mason 1995). In this section we wish to consider three issues raised in recent debates surrounding the evaluation of multilevel models that are of particular importance in geographical research.

Operationalizing Context

The first issue relates to operationalizing context in terms of the hierarchical structure of particular data sets (Mason 1995). For example, it is *HALS1*'s multistage sampling design that provides the structure for the present analysis. On practical grounds, this strategy makes perfect sense. This practical convenience may, however, be bought at a considerable cost in theoretical terms for at least three reasons. First, more often than not the structure derives from convenient administrative boundaries and while these may capture some notion of geographical context at different spatial scales, they often have no explicit justification in terms of the outcomes being studied. Second, such boundaries imply a functional carving up of space that does not rest easily with Massey's (1991, p. 277) argument that "localities are not simple areas you can easily draw a line around." Finally, if based only on broad spatial units there can be no recognition that "every citizen lies at the center of a social experience produced by a series of intersecting, overlapping, layered environments" (Huckfeldt, Plutzer, and Sprague 1993, p. 365). Jones and Duncan (1996) have identified one possible solution to these problems involving the use of Openshaw's (1977) automatic zoning procedure and multilevel models based on cross-classified structures (Goldstein 1994).

Unlike basic hierarchical structures where individuals nest within one and only one context at each higher-level, cross-classified structures allow individuals to nest within a number of overlapping, or crossed, contexts at each higher level. In fact, only recently have such structures become computationally tractable and work is now appearing that details how they can be specified and applied (Jones, Gould, and Watt 1998). One particularly good illustration of their substantive value is Goldstein's (1995) work in education where he reports a cross-classified multilevel model in which pupils are nested in primary schools that are crossed with secondary schools. Undertaking a simple pupil-within-secondary-school analysis while allowing for ability on intake suggests substantial differences in progress between secondary schools. The cross-classified analysis, however, shows that most of this difference is attributable to the primary school level. Thus, explanations for between-school differences need to be sought at more than the current school setting. In geographical research, the possibility of looking in the wrong place is ever present. Failing to recognize overlapping/multiple contexts can be regarded as a potential source of misinterpretation and misspecification in multilevel analysis. Cross-classified multilevel models, however, offer one important way of avoiding this.

Recent research has suggested that one other way of looking in the wrong place may come from ignoring the immediate family context, that is, the household. Interestingly, given our example, this has been prompted by research on drinking behavior in Britain. Using a different national survey from that used here, Rice et al. (1998) find that when households are incorporated as a level, between-local-neighborhood differences are extremely small. Obviously, this has implications for our findings and a number of both specific and more general points can be made. First, in our study households could not be included as a level since only one person per household was sampled in *HALS1*. Thus, due to the survey design, individuals and households were confounded. As Rice et al. (1998, p. 974) note, this is quite usual in large-scale surveys and emphasizes the need for survey design to be both practically and theoretically driven. Second, it should be noted that Rice et al.'s analysis considers only simple, constant between-place variation. Furthermore, at no point in the analysis are higher-level variables included. Given the findings here, these omissions may have as much significance for understanding drinking behavior as failing to take account of households. Third, there is no reason to believe that strong local geographies are always an artifact of ignoring households: multilevel research in other areas has, for example, found important variation between places *after* taking into account significant within-place, between-household variation (for example, Steele, Diamond, and Amin 1996; Pampalon et al. 1999).

The potential confounding effect of family factors has also been raised in recent years in a broader literature that addresses the study of contextual effects in relation to child and youth development [for a review see Duncan and Raudenbush (1999)]. Several workers have emphasized that since parents can, to some extent, choose where to live, unobservable family factors can easily be mistaken for neighborhood effects: as an example, parents who move to "better" neighborhoods might be those who spend more time helping their children with their homework (for example, Evans, Oates, and Schwab 1992). This issue—the endogeneity of contextual effects—is under-researched and little considered. Here we would like to outline briefly current thinking in this area while recognizing that debate on this complex issue is only now starting to develop.[16]

First, as Manski (1993) emphasizes, the only guaranteed solution to this problem is the use of experimental or quasi-experimental data. Following Duncan and Raudenbush (1999), however, this gold standard is not, itself, without other problems (for example, cost, representativeness, external validity). In view of this, a sensible approach would be not to dismiss evidence from observational designs out of hand but to work to ensure that such evidence is produced as carefully as possible. Presently, there are three main ways this can be achieved within a multilevel context. First, and most simply in analytical (though not survey) terms, is the inclusion of data that "measure the crucial omitted variables" (Duncan and Raudenbush 1999, p. 37). Second is the application of specially developed multilevel instrumental variable estimation techniques—one of the standard solutions to endogeneity problems in single-level regression (Spencer 1999). Third is the use of longitudinal fixed-effects models based on the nesting of panel observations for those who move neighborhoods within a cross-classified structure (Rasbash 2000). While the second and third options are technically complex and computationally intensive, they are the focus of ongoing research and are likely to become more tractable, and widely used, with time.

Finally, we would like to note that while solutions such as cross-classified structures certainly represent a new and important way of doing multilevel analysis, to us it does seem rather limiting to think that these problems reduce completely to a technical fix. Rather, we prefer to think of them as serving as a useful reminder that extensive, quantitative techniques such as multilevel modeling represent space and context and their effects in specific, particular ways (Dixon and Jones 1998).

Predictions and Uncertainty

The second issue arises from work focusing on the application of multilevel models in performance measurement. Notions of composition and context as described earlier in relation to geographical variations in individual outcomes have exact parallels in research examining variations in institutional performance. For example, variations in the performance of health service activities between different provider units (for example, vaccination uptake rates in clinics; length of stay times in hospitals) can be attributed to both the type of people particular units serve (compositional effects) and the nature of the environment from and in which the service is provided (contextual effects). To establish how well a particular unit is performing, adjustment must be made for the type of people it serves. One way of doing this is through multilevel analysis. In short, client-within-institution models need to be estimated including a range of individual-level variables describing patient characteristics. Predictions for specific institutions can then be obtained, conditional on patient composition, and these can be ranked to construct the now familiar league tables (Goldstein and Spiegelhalter 1996).

[16]This issue is unlikely to apply in the case of our empirical example in that people are unlikely to move to a particular area on the basis of factors relating to alcohol consumption.

Given this, it seems reasonable that the same procedure be used in geographical applications to identify unusual and anomalous areas or places. Recent writings on institutional performance, however, offer important guidance that researchers need to consider before doing this. First, it is apparent that there is, in some sense, a disjuncture between using multilevel models to generate place-specific predictions and their primary function of modeling population heterogeneity. As outlined earlier, since multilevel analysis treats the places explicitly identified as a sample from a population of places, the main focus is on the *variability between places* rather than the effects of specific named places. Immediately, therefore, there is some conceptual distance between multilevel techniques and their use in identifying particular anomalous instances.

In more practical terms, what needs to be remembered is that, while making predictions for specific places is possible based on shrinkage estimators, they are not simply point estimates; degrees of uncertainty are associated both with them and any rankings that derive from them (Goldstein and Spiegelhalter 1996). As research on educational performance is now showing, after careful modeling that adjusts for compositional factors and takes into account sample sizes, rankings carry such large uncertainty bands that distinguishing between the performance of most institutions is extremely difficult. For example, Draper (1995), commenting on work done by Goldstein and Thomas (1993) on school differences in pupil attainment, emphasizes that when uncertainty is taken into account the resulting categorization of performance is necessarily so broad that the majority of schools (70 percent) can be accurately located only "somewhere in the middle of a large gray area" (Draper 1995, p. 133). This is not a negative finding but is, instead, a positive, important result: schools might not be as different as we might think.

This and other examples emphasize that, while multilevel analysis provides a robust way of investigating place heterogeneity in the population, care must be taken when going beyond this to making specific statements about specific places. Crucially, place-specific estimates depend upon the sample size in specific places. Places with small sample sizes will have large confidence intervals; they will also contribute little to the estimation of the population parameters given the precision-weighting estimation procedure used. In brief, therefore, making comparisons between specific places requires reasonable within-place sample sizes [Paterson and Goldstein (1992) suggest a minimum of 25]. At the same time, there should also be as little difference as possible in sample sizes between places if detailed comparisons are to be made.

Exchangeability Judgments

Finally, it is important to consider further the point of disjuncture just mentioned in relation to making place-specific predictions. Multilevel models treat the higher level units as a sample drawn from a common population and inferences are made about this population. However, just because multilevel models operate in this way does not guarantee that they are appropriate in any particular instance. As Morris (1995, p. 198) writes, "there are crucial exchangeability judgments embraced (in multilevel models) that sometimes are not carefully considered in applications." Thus, before using the technique, researchers need to be sure that, first, the sample of places they are working with does, in general, come from (can be exchanged with/is similar to) the population that they wish to make inferences about and, second, that this holds true for each specific place for which they have data. If, for any reason, a researcher believes that certain places are truly separate or that they come from different populations, they should not be regarded as exchangeable with the remaining random sample of places and they need to be treated as fixed effects. In our study, for example, it might be that rural and urban places are so different that they comprise two separate populations in such a way that one should be modeled using a dummy/indicator-coded variable in the fixed part. Whilst checking for this situation can be done empirically after fitting a model and looking for outliers, there are dangers to such an approach as the shrinkage estimation procedure will give a misleading picture of similarity, especially when sample sizes between higher-level units are unbalanced. As Draper (1995) points out, the only way to guarantee exchangeability is by careful planning at the design stage before any data have been collected. Thus, researchers must ensure that their study design involves collecting data from a representative sample taken from a clearly defined target population.

Once again, therefore, there is a flexible solution within multilevel analysis: if, for whatever reason, our data does contain untypical places, or two or more different types of place, we can include appropriate dummy/indicator variables and interactions as fixed terms, while still allowing the remaining places to form a higher-level distribution. As before, however, this issue can also be thought of as prompting deeper, more philosophical questions. For us, it extends beyond issues of research design and takes us back to the heart of long-standing theoretical debates over generalization and exceptionalism (Bunge 1966; Johnston 1985).

CONCLUSIONS

Multilevel models have several features that make them attractive in geographically based research. In this paper, rather than concentrate on repeating their technical advantages, we have sought to explain and emphasize further how, in more general terms, they offer an extremely flexible, yet coherent framework for framing and evaluating ideas about contextuality and variability. In doing this, we hope to have shown that, as well as having the specific technical benefits well covered elsewhere in the literature, such models encourage and foster rigorous thinking about places and their effects.

As writers are increasingly recognizing, however, multilevel models are undoubtedly more complex, and, in the words of Draper (1995, p. 139), this opens "the possibility of interpretive confusion and overstatement of what may be validly concluded from a given body of evidence." This paper has highlighted several issues that we believe geographical researchers using these techniques need to be particularly aware of. Some of these are practical in nature and relate to model implementation and specification. Alternatively, they may be associated with the design features of certain software packages. Most importantly, the illustrative analysis emphasizes the key general point that quite different substantive interpretations are possible depending on which issues are considered and which are not. In the present instance, there may be grounds for thinking that local neighborhoods play more of a role in shaping drinking behavior, at least for certain types of people, than may immediately be apparent from simple models, be they single or multilevel.

These issues should not be seen as invalidating the use of multilevel models in geographical research. In fact, they signal quite the reverse for, as the analysis presented here shows, the flexibility and complexity of multilevel modeling can bring a range of new and important insights. In comparison, standard OLS regression modeling does seem over-reductionist, both in presuming that there are no differences between places and that between-individual differences can be captured by a single, summary measure. For these reasons, multilevel modeling techniques do, we believe, deserve to be widely used by geographers and others doing geographic-related research. At the same time, however, the present exposition should also serve as an important reminder that, like all statistical methods, multilevel models do need to be used "with care and understanding" (Goldstein 1995, p. 13).

LITERATURE CITED

Aitken, M. and N. Longford (1986). "Statistical Modeling Issues in School Effectiveness Studies." *Journal of the Royal Statistical Society (Series A)* 149, 1–26.

Best, N., D. Spiegelhalter, A. Thomas, and E. Bryne (1996). "Bayesian Analysis of Realistically Complex Models." *Journal of the Royal Statistical Society (Series A)* 159, 323–42.

Blaxter, M. (1990). *Health and Lifestyles*. London: Tavistock/Routledge.

Britton, M. (1990). *Mortality and Geography: A Review in the mid-1980s for England and Wales*. London: HMSO.

Bryk, A., and S. Raudenbush (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park: Sage.

Bryk, A., S. Raudenbush, M. Seltzer et al. (1986). *An Introduction to HLM: Computer Program and User's Guide*. Chicago Department of Education, University of Chicago.

Bullen, N., K. Jones, and C. Duncan (1997). "Modeling Complexity: Analyzing Between-Individual and Between-Place Variation—A Multilevel Tutorial." *Environment and Planning Series A* 29, 585–609.

Bunge, W. (1966). "Locations Are Not Unique." *Annals of the Association of American Geographers* 56, 375–76.

Carr-Hill, R., P. Smith, S. Martin, et al. (1994). "Allocating Resources to Health Authorities: Development of Methods for Small Area Analysis of Use of Inpatient Services." *British Medical Journal* 309, 1046–49.

Carmines, E., R. Huckfeldt, and C. McCurley (1995). "Mobilization, Counter-mobilization, and the Politics of Race." *Political Geography* 14, 601–19.

Casetti, E., and J. Jones, eds. (1992). *Applications of the Expansion Method*. London: Routledge.

Charnock, D. (1996). "National Uniformity and State and Local Effects on Australian Voting: A Multilevel Approach." *Australian Journal of Political Science* 31, 51–65.

Congdon, P. (1995). "The Impact of Area Context on Long-Term Illness and Premature Mortality: An Illustration of Multilevel Analysis." *Regional Studies* 29, 327–44.

Cox, B. (1988). *Health and Lifestyle Survey*, 1984–5 (computer file). Colchester: ESRC Data Archive.

Cox, B., M. Blaxter, A. Buckle, et al. (1987). *The Health and Lifestyle Survey: A Preliminary Report*. London: Health Promotion Research Trust.

Deegan, J. (1976). "The Consequences of Model Misspecification in Regression Analysis." *Multivariate Behavioral Research* 11, 237–48.

de Leeuw, J., and I. Kreft (1995). "Questioning Multilevel Models." *Journal of Educational and Behavioral Statistics* 20, 171–89.

DiezRoux, A. (1998). "Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis." *American Journal of Public Health* 88, 216–22.

DiPrete, T., and J. Forristal (1994). "Multilevel Models: Methods and Substance." *Annual Review of Sociology* 20, 331–57.

Dixon, D., and J. Jones (1998). "My Dinner with Derrida, Or Spatial Analysis and Poststructuralism Do Lunch." *Environment and Planning A* 30, 247–60.

Draper, D. (1995). "Inference and Hierarchical Modeling in the Social Sciences." *Journal of Educational and Behavioral Statistics* 20, 115–47.

Duncan, C., K. Jones, and G. Moon (1993). "Do Places Matter? A Multilevel Analysis of Regional Variations in Health-related Behavior in Britain." *Social Science and Medicine* 37, 725–33.

______ (1995). "'Psychiatric Morbidity: A Multilevel Approach to Regional Variations in the UK." *Journal of Epidemiology and Community Health* 49, 290–95.

______ (1999). "Smoking and Deprivation: Are There Neighbourhood Effects?" *Social Science and Medicine* 48, 497–505.

Duncan, G., and S. Raudenbush (1999). "Assessing the Effects of Context in Studies of Child and Youth Development." *Educational Psychologist* 34, 29–41.

Evans, W., W. Oates, and R. Schwab (1992). "Measuring Peer Group Effects: A Study of Teenage Behavior." *Journal of Political Economy* 100, 966–91.

Gilks, W., D. Clayton, D. Spiegelhalter, et al. (1993). "Modeling Complexity: Applications of Gibbs Sampling in Medicine (with discussion)." *Journal of the Royal Statistical Society Series B* 55, 39–102.

Goldstein, H. (1994). "Multilevel Cross-classified Models." *Sociological Methods and Research* 22, 364–75.

_____ (1995). *Multilevel Statistical Models*. London: Edward Arnold.

Goldstein, H., J. Rasbash, I. Plewis, et al. (1998). *A User's Guide to MLwiN*. London: Institute of Education, University of London.

Goldstein, H., and D. Spiegelhalter (1996). "League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance (with discussion)." *Journal of the Royal Statistical Society Series A* 159, 385–443.

Goldstein, H., and S. Thomas (1993). *Guardian A-Level Analysis 1993: Technical Report*. London: Institute of Education, University of London.

Gould, M., and K. Jones (1996). "Analyzing Perceived Limiting Long-Term Illness Using U.K. Census Microdata." *Social Science and Medicine* 42, 857–69.

Hauser, R. (1970). "Context and Consex: A Cautionary Tale." *American Journal of Sociology* 75, 645–64.

Hedeker, D., S. McMahon, L. Jason, and D. Salina (1994). "Analysis of Clustered Data in Community Psychology: With an Example from a Worksite Smoking Cessation Project." *American Journal of Community Epidemiology* 22, 595–615.

Hox, J., and I. Kreft (1994). "Multilevel Analysis Methods." *Sociological Methods and Research* 22, 283–99.

Huckfeldt, R., E. Plutzer, and J. Sprague (1993). "Alternative Contexts of Political Behavior: Churches, Neighbourhoods and Individuals." *The Journal of Politics* 55, 365–81.

Humphreys, K., and R. Carr-Hill (1991). "Area Variations in Health Outcomes: Artefact or Ecology." *International Journal of Epidemiology* 20, 251–58.

Johnston, R. (1985). "To the Ends of the Earth." In *The Future of Geography,* edited by R. Johnston, pp. 326–38. London: Methuen.

Johnston, R., C. Pattie, and J. Allsopp (1988). *A Nation Dividing?* London: Longman.

Jones, K. (1991). "Specifying and Estimating Multilevel Models for Geographical Research." *Transactions of the Institute of British Geographers* 16, 148–59.

Jones, K., and N. Bullen (1993). "A Multilevel Analysis of the Variations in Domestic Property Prices: Southern England, 1980–1987." *Urban Studies* 30, 1409–26.

Jones, K., and N. Bullen (1994). "Contextual Models of House Prices: A Comparison of Fixed- and Random-Coefficient Models Developed by Expansion." *Economic Geography* 70, 252–72.

Jones, K., and C. Duncan (1996). "People and Places: The Multilevel Model as a General Framework for the Quantitative Analysis of Geographical Data." In *Spatial Analysis: Modeling in a GIS Environment,* edited by P. Longley and M. Batty, pp. 79–104. Cambridge: Longman.

Jones, K., M. Gould, and R.Watt (1998). "Multiple Contexts as Cross-classified Models: The Labor Vote in the British General Election of 1992." *Geographical Analysis* 30, 65–93.

Jones, K., R. Johnston, and C. Pattie (1992). "People, Places and Regions: Exploring the Use of Multilevel Modeling in the Analysis of Electoral Data." *British Journal of Political Science* 22, 343–80.

Kreft, I., J. de Leeuw, and R. Van der Leeden (1994). "Review of Five Multilevel Analysis Programs: BMDP-5V, GENMOD, HLM, ML3, VARCL." *The American Statistician* 48, 324–35.

Langford, I., and G. Bentham (1997). "A Multilevel Model of Sudden Infant Death Syndrome in England and Wales." *Environment and Planning A* 29, 629–40.

Lewis, G., and M. Booth (1994). "Are Cities Bad for Your Mental Health." *Psychological Medicine* 24, 913–15.

Macintyre, S., S. MacIver, and A. Sooman (1993). "Area, Class and Health: Should We Be Focusing on People or Places?" *Journal of Social Policy* 22, 213–34.

Manski, C. (1993). "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60, 531–42.

Mason, W. (1995). "Comment." *Journal of Educational and Behavioral Statistics* 20, 221–27.

Massey, D. (1991). "The Political Place of Locality Studies." *Environment and Planning Series A* 23, 267–81.

Mitchell, R., S. Gleave, K. Lynch, et al. (1998). "Health: Where You Are or Where You Live." *Health Variations—Issue Two,* Lancaster: ESRC.

Morris, C. (1983). "Parametric Empirical Bayes." *Journal of the American Statistical Association* 78, 47–65.

_____ (1995). "Hierarchical Models for Educational Data: An Overview." *Journal of Educational and Behavioral Statistics* 20, 190–99.

Openshaw, S. (1977). "A Geographical Solution to Scale and Aggregate Problems in Region-building, Partitioning and Spatial Modeling." *Transactions of British Geographers* NS2, 459–72.

Pampalon, R., C. Duncan, S. Subramanian, and K. Jones (1998). "Geographies of Health Perception in Quebec: A Multilevel Perspective." *Social Science and Medicine* 48, 1483–90.

Paterson, L., and H. Goldstein (1992). "New Statistical Methods for Analyzing Social Structures: An Introduction to Multilevel Models." *British Educational Research Journal* 17, 387–93.

Rasbash, J. (2000). "Re: Participants Who Move." http://www.mailbase.ac.uk/lists/multilevel/2000-01/0062.html (accessed 23 February 2000).

Raudenbush, S., and J. Willms (1991). *Schools, Classrooms and Pupils*. San Diego: Academic Press.

Reijneveld, S. (1998). "The Impact of Individual and Area Characteristics on Urban Socioeconomic Differences in Health and Smoking." *International Journal of Epidemiology* 27, 33–40.

Rice, N., R. Carr-Hill, P. Dixon, and M. Sutton (1998). "The Influence of Households on Drinking Behavior: A Multilevel Analysis." *Social Science and Medicine* 46, 971–79.

Shouls, S., P. Congdon, and S. Curtis (1996). "Modeling Inequality in Reported Long-Term Illness in the UK: Combining Individual and Area Characteristics." *Journal of Epidemiology and Community Health* 50, 366–76.

Silk, J. (1977). "Analysis of Covariance and Comparison of Regression Lines." *Concepts and Techniques in Modern Geography 20*. Norwich: Geobooks.

Snijders, T., and R. Bosker (1999). "Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling." London: Sage.

Spencer, N. (1999). "Re: Lagged-Covariate." http://www.mailbase.ac.uk/lists/multilevel/1999–06/0000.html (accessed 23 February 2000).

Steele, F., I. Diamond, and S. Amin (1996). "Immunization Uptake in Rural Bangladesh: A Multilevel Analysis." *Journal of the Royal Statistical Society Series A* 159, 289–99.

Verheij, R., D. de Bakker, and P. Groenewegen (1999). "Is There a Geography of Alternative Medical Treatment in The Netherlands?" *Health and Place* 5, 83–97.

Von Korff, M., T. Koepsell, S. Curry, et al. (1992) "Multi-level Analysis in Epidemiologic Research on Health Behaviors and Outcomes." *American Journal of Epidemiology* 135, 1077–82.

Weisberg, S. (1980) *Applied Linear Regression*. New York: Wiley.

Woodhouse, G., ed. (1995). *A Guide to MLn for New Users*. London: Institute of Education, University of London.