Published in final edited form as:

Epidemiology. 2010 July ; 21(4): 475–481. doi:10.1097/EDE.0b013e3181d74a71.

Modeling Neighborhood Effects:

The Futility of Comparing Mixed and Marginal Approaches

S. V. Subramanian^a and A. James O'Malley^b

^aDepartment of Society, Human Development and Health, Harvard School of Public Health, Boston, MA

^bDepartment of Health Care Policy, Harvard Medical School, Boston, MA

Using simulation techniques, Hubbard et al¹ compare mixed (also known as multilevel or hierarchical)^{2,3} and marginal (also known as population average or generalized estimation equation [GEE])⁴ approaches to modeling neighborhood effects. The choice of modeling approach has not received much prominence in social epidemiology. Hubbard et al are to be congratulated for providing a concise statistical summary outlining the differences between mixed and marginal models, especially with regard to model interpretation. The article also offers useful insight into several compelling statistical issues. Hubbard et al rightly recommend that researchers consider a wide range of issues before choosing one modeling strategy over the other.

It is their assertion of the superiority of the marginal approach to modeling neighborhood effects that stimulates our commentary. We argue that comparing 2 modeling approaches with fundamentally distinct targets of inference is futile. Indeed, researchers should decide on the modeling approach after having chosen the question of interest, and not the other way around. We start by discussing whether the selection of modeling approaches for neighborhood effects should be based on conceptual or statistical grounds. We then argue that differences in the modeling approaches are likely to be inconsequential for estimation of the marginal effect of a neighborhood attribute. We then review the role of parametric models in deriving GEE, and provide one suggestion for reducing the sensitivity of results to the random effects distribution in mixed models. We consider the implications of nonlinearity with respect to parameters and argue for mapping effects to the scale that practitioners find the easiest to interpret. Finally, we comment on the target of inference and the relative merits of descriptive and generative modeling.

CHOICE OF MODELING APPROACH: STATISTICAL OR SUBSTANTIVE?

Hubbard et al¹ favor the marginal approach because it involves fewer assumptions and does not require knowledge of the ways in which individuals are correlated within a neighborhood nor on how neighborhoods vary. Yet, without having to tax our brains to understand what may have generated the data structure in terms of correlations and variances, marginal approaches have a unique capability of providing the “right” answer, albeit (and somewhat unfortunately) to one question, which is, on average is there an association between 2 variables? As Raudenbush has cautioned, “that the marginal answer is robust does not make it a better answer unless the scientific question truly requires a marginal inference.”⁵

Showing the average association between a neighborhood attribute and individual health is only one of several ways to consider neighborhood effects. For instance, when an association between an individual outcome and individual predictor varies across neighborhoods, should that not be considered a neighborhood effect? One may also wish to determine whether the average effect of a neighborhood attribute varies across different cities or regions, in addition to reporting the “global” average effect? Furthermore, what if we want to know the conditional effect of living in a specific neighborhood, with varying samples of individuals in them? In short, the fundamental question is whether the complex heterogeneities underlying the marginal associations are simply a nuisance, or are they equally important for interpreting neighborhood effects.^{6,7} It would seem highly unrealistic, especially for social epidemiologic research, to disregard heterogeneity in the association while drawing substantive inferences.

It is well known that regression coefficients in mixed models are more sensitive to the assumption of random effects than their counterparts in marginal formulations.^{3,8,9} The basic concern is that the assumption of random effects in a mixed model is not verifiable. Consequently, a marginal formulation, using GEE, is often argued (including by Hubbard et al¹), to be a preferred method. Such a justification, as Raudenbush points out, is logically problematic.⁵ If mixed-model results vary as a function of assumptions about random effects, at least some of them must also differ from marginal results regarding the apparent association between predictor and outcome.⁵ Situations surely exist in which the conditional result from the “true” mixed model leads to a conclusion different from the associated marginal result. Consequently, why worry about the comparative robustness of mixed and marginal approaches, when they inevitably differ because they estimate different quantities? These are not “philosophical considerations,” as Hubbard et al seem to imply.¹ Marginal approaches are fundamentally incapable of answering complex, and arguably more interesting and relevant, questions that are pertinent for a comprehensive understanding of how and why neighborhoods might matter for an individual's health.

MARGINAL VERSUS MIXED: DOES IT MATTER, PRACTICALLY SPEAKING?

Let us assume that in the field of neighborhoods and health researchers have favored mixed over marginal approaches when the data structure is multilevel (ie, an outcome is measured on individuals at level 1, individual observations are nested within neighborhood at level 2 with an attribute measured at the neighborhood level).⁷ This preference could well be because mixed models provide answers to a broader range of neighborhood-related questions than does a marginal approach, making mixed models a substantive and logical choice. But, what about situations in which researchers are interested only in the marginal association between a neighborhood attribute and individual health outcome? Have studies using mixed-models gotten the wrong answer to this question? Hubbard et al¹ imply that empirical answers derived from mixed models on marginal associations between a neighborhood attribute and individual health are flawed. Unfortunately, because Hubbard et al do not apply their comparative statistical framework to an empirical dataset, we cannot assess whether published inferences from mixed models on the marginal associations are off the mark.

It is worth situating the interest in comparing mixed and marginal approaches in its historical context. Although concerns related to marginal and mixed models may not have caught the attention of social epidemiologists (with few exceptions),¹⁰ they have been central to discussions in applied longitudinal analysis.^{3,11,12} The data structure in applied longitudinal analysis is a special case of multilevel structure, with repeated measurements over time at level 1 nested within the same individual at level 2.^{6,13} Indeed, the development

of mixed models,^{14,15} as well as marginal models using GEE,³ had its origin in biostatistics, with the mixed approaches preceding the marginal. In longitudinal data settings, where random effects at the individual level were large (ie, the correlation between repeated measures within individuals were substantially higher leading to greater differences between individuals), sensitivity of the regression coefficients to random effects was a major concern. Importantly, there was also little interest in modeling the individual growth trajectory in outcomes over time, which mixed models could efficiently handle. Simply put, the correlation among repeated measures for an individual was largely considered a nuisance parameter. Thus, even though mixed models offered a comprehensive approach to accounting for the correlation in the data, these models were too cumbersome and demanding, especially since the inferential goal was marginal. The marginal approach, using GEE, in this specific context provided an elegant alternative that was simpler and robust enabling marginal inference that accounted for correlation among repeated measures within the individual. Thus, in applied longitudinal research adoption of the marginal models made intuitive sense because the target of inference was not the individual, but population; and when mixed models were used, given the high similarity within individuals leading to large random effects, the regression coefficient (the only parameter of interest) might be a very different quantity than the marginal effect.

The context of neighborhoods and health research is markedly different. Similarity among individuals within a neighborhood is often of substantive interest (ie, generated due to exposure to a shared environment) and not a nuisance. Furthermore, specific neighborhoods are also a target of inference (as in small area estimation),¹⁶ making mixed models a natural candidate for analysis. At the same time, intraclass correlations in neighborhoods (ie, the extent to which individuals are similar within a neighborhood) are typically considerably smaller than those observed in longitudinal analysis. In nonlinear models applied to neighborhood and health data, intraclass correlations are typically less than 2%, and occasionally between 3% and 5%. In linear models, stronger intraclass correlations have been observed (~10%). But with linear models the marginal and mixed approaches are equivalent, and the whole issue of marginal versus mixed is moot from a statistical standpoint. Thus, the implications of the technical criticisms of mixed-models presented by Hubbard et al for the body of work that has used mixed-modeling approaches may be inconsequential from a practical perspective. Indeed, there was a missed opportunity: Hubbard et al¹ could have tested their statistical claims for neighborhoods and health research and answered an important question: are published neighborhood-health associations sensitive to modeling choices?

We hasten to add that the issue is not simply one of intraclass correlations. If neighborhoods account for only a small fraction of the total variance, some may ask, why focus on them? Interpretations based solely on summary statistics such as intraclass correlation (defined as the proportion of neighborhood variance divided by the total variance in mixed models) is problematic because it implicitly assumes that the “systematic” and “stochastic” components of the variation within each of the individual and neighborhood levels is exactly the same. This is unlikely to be the case; at the individual level the stochastic component of the variation is likely to be very high, while the opposite is true at the neighborhood level. Thus, the focus ought to be more on the size of the random effects (ie, variance at neighborhood level) as opposed to the relative contribution of neighborhood to total variation.

INVOLVEMENT OF PARAMETRIC ASSUMPTIONS

Hubbard et al¹ overlook an important commonality between marginal and mixed models, which is that both involve parametric assumptions. Although marginal modeling via GEE is not invoked by a specific parametric assumption, the approach emanates from the likelihood

equations for generalized linear models (GLMs) and therefore is a quasi-likelihood procedure.^{8,9} Furthermore, the adaptation of the likelihood equations for GLMs to obtain GEE introduces a correlation/association matrix analogous to the way such a matrix extends the normal equations for linear regression to those for a random intercept mixed effects linear model. Thus, a marginal model is in essence linked to the distribution of the data in a more direct way than pure nonparametric procedures, such as rank-based methods.

It is important to realize that theory for marginal models is built on parametric assumptions. It is also important to recognize that mixed models are not restricted to solely parametric specifications of the random-effects distribution. For example, in Bayesian modeling a Dirichlet process prior is a semi-parametric alternative to normally distributed random effects.^{17–19} A good feature of this alternative is that it lets the data determine the distribution for random effects, thereby reducing reliance of the model on parametric assumptions, and making inferences based on the mixed model more robust.

SCALE ON WHICH EFFECTS ARE REPORTED

Hubbard et al imply that, due to the equivalence of conditional and marginal effects, mixed modeling is more attractive in the linear case than the nonlinear case. This distinction is troublesome as it suggests that one approach might be best for one type of data, and with the other preferred for other types. One could reasonably expect that the strength of argument for mixed models over marginal models (or vice versa) should be invariant to the type of data.

Quantitative differences between results for mixed and marginal models can be sensitive to the scale on which they are reported. In practice, it is useful to map results onto the scale of the outcome (eg, the [0, 1] interval in the case of a binary outcome), obtaining “common-language effect sizes.” Indeed, substantive researchers and other practitioners typically find it easiest to interpret effects on the same scale as the outcome.²⁰ Although the required computations are more challenging in the nonlinear case, modern methods for statistical computation can routinely handle the evaluation of expectations of nonlinear functions of random variables. In particular, for Bayesian nonlinear models fit using Markov-chain Monte-Carlo methods such as the Gibbs sampler,²¹ estimates of marginal effects on the scale of the data can be computed just as easily as their conditional counterparts; parameter values are drawn from the joint posterior distribution and then evaluated as Monte-Carlo averages of the (conditional or marginal) quantity of interest.^{22,23}

TARGET OF INFERENCE

Hubbard et al¹ argue that existence of “an interpretable parameter that can be defined as the projection of a misspecified mixed model onto the true underlying model”⁴ is a reason for choosing one approach over the other. We disagree. Basing an analysis on a particular population quantity of interest (eg, a linear trend) as opposed to modeling the data-generating process, which might imply a different relationship, implies the target of inference involves only the sampled or observed neighborhoods. In this finite-population scenario, fixed-effect specifications (marginal models) are appropriate. However, in many (if not most) situations, the actual population of interest is the “superpopulation” of all neighborhoods and residents, in which case the sampled neighborhoods and sampled individuals should be thought of as being one realization of the superpopulation of neighborhoods and individuals. After fitting a generative (ie, mixed) model, results can be depicted by approximating the estimated conditional or marginal relationship between a predictor and the outcome by a descriptive model, such as a linear trend.

SUMMARY

In summary, Hubbard et al¹ make an important contribution to the social epidemiological literature by providing an excellent introduction to marginal and mixed approaches, summarizing the interpretative differences, and raising a number of technical points. Although such comparisons could be made purely on technical grounds, the substantive and empirical context is also a key consideration that researchers should not ignore while making modeling choices. Marginal models are incapable of accommodating a richer set of questions that are pertinent for neighborhood and health research, and for which mixed models constitute a more appropriate conceptual and empirical choice. Even from a purely statistical perspective, it is clear that one can derive marginal estimates from a mixed model in a semi-parametric manner, but the reverse is not possible. It is worth emphasizing that a complex model (of the data generating process) can tell us about simpler specifications (as might be of interest for descriptive purposes), but a simpler model can never provide insight about the complex model.

Acknowledgments

We thank James Ware and Jarvis Chen for helpful discussions on the subject of mixed and marginal models.

Supported by the National Institutes of Health Career Development Award (NHLBI K25 HL081275) (to S.V.S.).

Biographies

S. V. SUBRAMANIAN is an associate professor at the Harvard School of Public Health. His research is on multilevel approaches to understanding the determinants of population health. His current work focuses on ways in which place and social context influence individual health outcomes.

A. JAMES O'MALLEY is an associate professor of statistics at the Harvard Medical School. His interests include Bayesian statistics, social network analysis, multivariate hierarchical models and causal inference in health research. His applied research focuses on the relationship of health and social networks, measurement of quality, and long-term care.

REFERENCES

1. Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*. 2010; 21:467–474. [PubMed: 20220526]
2. Goldstein, H. *Multilevel Statistical Models*. London: Arnold; 2003.
3. Raudenbush, SW.; Bryk, AS. *Advanced Quantitative Techniques in the Social Sciences 1*. 2nd ed.. Thousand Oaks, CA: Sage Publications; 2002. *Hierarchical linear models: Applications and data analysis methods*.
4. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988; 44:1049–1060. [PubMed: 3233245]
5. Raudenbush S. Marginalized multilevel models and likelihood inference: Comment. *Stat Sci*. 2000; 15:22–24.
6. Subramanian SV. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Soc Sci Med*. 2004; 58:1961–1967. [PubMed: 15020011]
7. Subramanian SV, Jones K, Kaddour A, Krieger N. Revisiting Robinson: the perils of individualistic and ecologic fallacy. *Int J Epidemiol*. 2009; 38:342–360. author reply 370–373. [PubMed: 19179348]
8. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986; 42:121–130. [PubMed: 3719049]

9. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
10. Subramanian SV, Kawachi I. Income inequality and health: what have we learned so far? *Epidemiol Rev*. 2004; 26:78–91. [PubMed: 15234949]
11. Fitzmaurice, G.; Laird, N.; Ware, JH. *Applied Longitudinal Analysis*. Chichester, United Kingdom: John Wiley; 2004.
12. Heagerty PJ, Zeger SL. Marginalized multilevel models and likelihood inference (with discussion). *Stat Sci*. 2000; 15:1–26.
13. Subramanian, SV.; Jones, K.; Duncan, C. Multilevel methods for public health research. In: Kawachi, I.; Berkman, LF., editors. *Neighborhoods and Health*. New York: Oxford University Press; 2003. p. 65-111.
14. Laird N, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
15. Henderson CR, Searle SR, Schaeffer LR. The invariance and calculation of method 2 for estimating variance components. *Biometrics*. 1974; 30:583–588.
16. Malec D. Small area estimation from the American community survey using a hierarchical logistic model of persons and housing units. *J Off Stat*. 2005; 21:411–432.
17. Zhang L, Mukherjee B, Hu B, Moreno V, Cooney KA. Semiparametric Bayesian modeling of random genetic effects in family-based association studies. *Stat Med*. 2009; 28:113–139. [PubMed: 18792083]
18. Dorazio RM, Mukherjee B, Zhang L, Ghosh M, Jelks HL, Jordan F. Modeling unobserved sources of heterogeneity in animal abundance using a Dirichlet process prior. *Biometrics*. 2008; 64:635–644. [PubMed: 17680831]
19. Ghosh P, Gonen M. Bayesian modeling of multivariate average bioequivalence. *Stat Med*. 2008; 27:2402–2419. [PubMed: 18095275]
20. McGraw KO, Wong SP. A common language effect size statistic. *Psychol Bull*. 1992; 111:361–365.
21. Casella G, George EI. Explaining the Gibbs sampler. *Am Stat*. 1992; 46:167–174.
22. Wu L. Exact and approximate inferences for nonlinear mixed-effects models with missing covariates. *J Am Stat Assoc*. 2004; 99:700–709.
23. Fotouhi AR. Comparisons of estimation procedures for nonlinear multilevel models. *J Stat Softw*. 2003; 8:1–39.