Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2006 : Methodological Issues in Measuring Population Health

2006

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2006
Methodological Issues in Measuring Population Health

Data Sparseness in Contextual Population Health Research: Effects of Small Group Size and Cluster Analysis on Linear and Non-Linear Multilevel Models

Philippa Clarke[1], Patricia O'Campo[2], and Blair Wheaton[3]

Abstract

The current use of multilevel models to examine the effects of surrounding contexts on health outcomes attest to their value as a statistical method for analyzing grouped data. But the use of multilevel modeling with data from population-based surveys is often limited by the small number of cases per level-2 unit, prompting a recent trend in the neighborhood literature to apply cluster analysis techniques to address the problem of data sparseness. In this paper we use Monte Carlo simulations to investigate the effects of marginal group sizes and cluster analysis techniques on the validity of parameter estimates in both linear and non-linear multilevel models.

KEY WORDS: multilevel models, data sparseness, cluster analysis, Monte Carlo simulations, survey research

1. Introduction

1.1 Data Sparseness in Multilevel Models

Population health outcomes are shaped by complex interactions between individuals and the diverse social, cultural, and environmental contexts in which they are situated over the life course. Physical and social environments vary across neighborhoods, schools, regions, and nations. These divergent contexts in which persons are situated contribute to population health disparities, as mortality, disease prevalence, physical function and maternal health vary across contexts (Yen & Syme, 1999; O'Campo, et al., 1997; Yen & Kaplan, 1999; Barr et al., 2001; Clarke & George, 2005). The recent increase in the use of multilevel models to examine associations between these group-level contexts and a wide range of individual health indicators (e.g. Pickett and Pearl 2001; Ahern, Pickett & Selvin, 2003; Buka et al., 2003; Merlo, et al., 2003; Ross & Mirowsky, 2001) attest to their value as a statistical method for analyzing grouped or clustered data. But the use of multilevel modeling with data from population-based surveys is often limited by *data sparseness*: a small number of cases per level-2 unit.

While large scale surveys make it relatively easy to achieve a large number of groups, there are often very few individuals per group. For example, in Cycle 1.1 of the Canadian Community Health Survey there are on average about 15 respondents per census tract (a typical geographic area used to approximate neighborhoods). The matter is further complicated by the fact that 5.9% of the tracts have only one respondent. These "singleton" groups generate concern because there is no within group variability, rendering individual- and tract-level variation indistinguishable. Sparseness is also evident in American survey data. The National Longitudinal Study of Adolescent Health is a United States school-based survey of health behaviors in adolescents and their outcomes in young adulthood (Bearman, Jones & Udry, 2002). However, because schools (not neighborhoods) were the sampling frame, there is considerable sparseness in the data for those interested in examining neighborhood effects on health. In the first wave of the survey there is an average of 7.33 subjects per tract, and almost half of these tracts are singleton tracts. Due to subject attrition and residential mobility, sparseness only increases over surveys with longitudinal follow-up.

---

[1] Philippa Clarke, University of Michigan, 426 Thompson Street, Ann Arbor, Michigan, 48106, USA (pjclarke@umich.edu);
[2] Patricia O'Campo, University of Toronto, 70 Richmond St. E., 4th Floor, Toronto, Canada; Blair
[3] Wheaton, University of Toronto, 725 Spadina Ave., Toronto, Canada.

When there is a high level of clustering within groups (i.e. large group size), it is well known that disaggregation of the data (e.g. by using ordinary least squares (OLS) regression) leads to an elevated risk of Type I error when examining the effect of group contexts on health (Kreft, 1996). By pretending that the observations are independent, the standard errors are biased downwards generating artificially narrow confidence intervals. Multilevel models appropriately partition within-group and between-group effects so that a high level of clustering within groups is statistically accounted for. But little is known about the *lower* threshold at which data sparseness renders multilevel models unreliable or even unnecessary. There are various rules-of-thumb stated in the literature, often in the range of 15 to 30 per group (Bryk and Raudenbush 1992; Kreft and de Leeuw 1998; Raudenbush and Sampson 1999). However, rules of thumb may be cited and followed without real proof of what the minimum level of data sparseness really is. And the wide range of numbers given in these recommendations reflects the fact that very little research has been done to explicitly test the minimum level of clustering necessary for valid and reliable estimates in multilevel models.

Simulations designed to assess the data sparseness problem are beginning to appear in the literature, and results suggest that the number of groups is more important for unbiased and efficient estimates than the number of observations per group (Maas and Hox 2002; Maas and Hox 2004; Afshartous 1995; Kreft 1996; Mok, 1995). This is reassuring since a shortage of groups is rarely a problem in population-based survey data. Yet researchers continue to be concerned with small group sizes when examining contextual effects on health and have adopted various strategies to deal with data sparseness. Some choose to ignore the hierarchical structure of the data altogether by using OLS regression techniques (e.g. Robert, 1998; South & Baumer, 2000; Schieman, Pearlin & Meersman, 2006), while others have simply deleted sparse neighborhoods from their analyses (Cutrona et al., 2000). There has also been a recent trend in the use of cluster analysis techniques to reduce data sparseness (Beland, Birch, and Stoddart 2002; Buka, et al. 2003; Cutrona, et al. 2000; Hou and Chen 2003; Wheaton & Clarke, 2003). This typically involves the use of a clustering procedure where respondents are grouped together into larger "synthetic" neighborhoods, yielding a larger number of cases per level-2 unit. Yet, the effects of such techniques on the accuracy of model parameters have not been thoroughly considered. In earlier work (Clarke & Wheaton, 2007) we showed that aggressive clustering strategies have the ironic consequence of actually minimizing between-group variance as a result of introducing artificial within-group heterogeneity. Targeted clustering of the singleton groups was the most effective strategy for reducing data sparseness while minimizing the inflation of within-group variance (Clarke & Wheaton, 2007; Wheaton & Clarke, 2003). While this work focused on linear hierarchical models, the effect of clustering on non-linear models remains unknown.

Our purpose in this paper is to empirically examine the effects of data sparseness on multilevel models so that informed analytic decisions can be made when working with grouped or clustered data. Using Monte Carlo simulations, we investigate the effects of marginal group sizes on parameter estimates in the multilevel model. We look at both linear and nonlinear hierarchical models to examine potential differences in the effects of marginal group size on both continuous and discrete outcomes. We then employ cluster analysis techniques to minimize data sparseness and examine the consequences for both models in the simulations. We argue that more sparseness may be tolerated than is generally assumed, and that any corrective actions should be used cautiously.

1.2 **The Generalized Multilevel Model**

In general, the multilevel model can be conceptualized as a hierarchical system of regression equations within J contextual groups, with N_j individuals in each group (Raudenbush and Bryk 2002). At the individual level (level-1) we have the dependent variable Y_{ij} and there are separate regression equations for each group or "neighborhood". For, linear models the identity link function regresses Y_{ij} on a linear predictor set of one (or more) independent variables X_{ij}, with normally distributed residuals (e_{ij}) having a mean of 0 and variance σ^2 :

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij} \qquad (1a).$$

For non-linear models, various link functions linearize an underlying non-linear predictor component. For the case of a binary outcome with a binomial error distribution, the logit link function is used to regress the log odds of Y_{ij} on a linear predictor set of independent variables:

$$Logit(Y_{ij}) = ln\left(\frac{P_{ij}}{1-P_{ij}}\right) = \beta_{0j} + \beta_{1j} X_{ij} \qquad (1b).$$

For both the linear and non-linear models these level-1 coefficients can then be modeled by explanatory variables at the contextual level-2 (e.g. neighborhood poverty):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j} \quad (3).$$

By substituting equations (2) and (3) into equations (1a) and (1b) and rearranging terms, we get the full multilevel linear model:

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{0j} + u_{1j} X_{ij} + e_{ij} \quad (4a),$$

and the full multilevel logistic model:

$$\log it(Y_{ij}) = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{0j} + u_{1j} X_{ij} \quad (4b).$$

In both models, u_{0j} represents group level variability around the intercept, which is assumed to be normally distributed with a mean of 0 and variance τ_{00}, and u_{1j} represents group level variability around the regression slope, which is assumed to be normally distributed with mean of 0 and variance τ_{11}.

If there are no explanatory variables at levels 1 or 2, equations (4a) and (4b) reduce to:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (5a),$$

$$\log it(Y_{ij}) = \gamma_{00} + u_{0j} \quad (5b),$$

which are the fully unconditional, or one-way ANOVA, models for the linear and logistic case, respectively. Partitioning the variance components yields a useful statistic, the intraclass correlation coefficient (ICC), which measures the proportion of variance in the outcome that is accounted for by the group level (Raudenbush and Bryk 2002). For the linear model the ICC is defined as:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \quad (6a).$$

Since the binomial distribution for the logistic link function implies a level 1 variance of $\pi^2/3$ (Snijders & Bosker, 1999), the ICC for the non-linear model is defined as:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \pi^2/3}. \quad (6b)$$

2. Methods

2.1 Simulation Procedure

A Monte Carlo simulation is conducted with a two-level hierarchical model. The number of groups is held constant at 200 groups, and group size is varied at 2, 5, 10, and 20 observations per group. The number of groups is chosen to represent the larger number of groups typically found in population-based survey data, while the group sizes capture the extremes of data sparseness as well as the larger group sizes typically tested in simulations (Mok 1995; Maas & Hox 2004). For each of these 4 conditions 1000 simulated data sets are generated for both the linear and nonlinear models.

The parameters for the simulation were set according to the full multilevel models (Eq. 4a and 4b). The intercept (γ_{00}) is set to 1.00, and the fixed effect coefficients ($\gamma_{10}, \gamma_{01}, \gamma_{11}$) to .3, representing a medium effect size (Cohen 1988). A set of X and Z values are randomly generated from a standard uniform distribution. Following Maas and Hox (2004) the residual variance at level-1 (σ^2) is fixed to .5 in the linear model. (There is no level-1 error in the non-linear model.)

The population values of the level-2 variance components are derived according to the formula for an ICC value of .1 (capturing the lower threshold of group level clustering on the outcome variable typically seen in population-based survey data (Gulliford et al., 1999)). Thus, for the nonlinear model the population value for the level-2 intercept variance (τ_{00}) is set to .366, based on equation 6b. For simplicity, τ_{00} and τ_{11} are constrained to be equal (following Maas and Hox 2004) and the level-2 covariance (τ_{01}) is set to zero. For the linear model equation 6a is modified to take into account heteroscedasticity in the random error term u_{1j} (where it is a function of the level-1 independent variable (X_{ij})) (Goldstein 1995; Mok 1995):

$$\rho = \frac{\text{var}(level-2)}{\text{var}(level-2) + \text{var}(level-1)} \tag{7a),}$$

where $$\text{var}(level-2) = \tau_{00} + 2\tau_{01}x_{ij} + \tau_{11}x_{ij}^2, \tag{7b)}$$

and $$\text{var}(level-1) = \sigma^2 \tag{7c).}$$

Thus, for the linear multilevel model the population variance at level-2 is set to .044 (based on an average x_{ij} value of .5 generated randomly from a uniform distribution). (No equivalent adjusted ICC formula is available for nonlinear models, so only the residual ICC formula is used.)

Based on these parameters, values of Y_{ij} are generated for each of the 4,000 simulated data sets (Y is continuous for the linear model, while for the non linear model Y can take values of 0 or 1), and the effects of the 4 different conditions on the estimated parameter values are examined for both the linear and non-linear model. Reproducible streams of random numbers were generated in the simulations in order to maintain comparability across models. All simulations are conducted in Mplus Version 4.2 (Muthen and Muthen, 1998). Models are estimated using maximum likelihood and the standard errors are computed using a robust sandwich estimator. Nonlinear models are estimated using a numerical integration algorithm.

2.2 Cluster Analysis

After generating the simulated data for each condition, we then performed a disjoint cluster analysis to collect groups together into larger "synthetic neighborhoods" with similar characteristics. Clustering was based on the similarity of groups according to the group-level variables that were generated in the initial simulated data. We used the FASTCLUS procedure in SAS to assign observations to one and only one cluster. (Clusters do not form a tree structure as they do in a hierarchical cluster analysis (Anderberg 1973).) Observations were grouped together to form a cluster based on Euclidean distances between values on the group-level variables. We used a strict distance criterion to ensure a high degree of similarity before groups could be clustered together into synthetic neighborhoods.

2.3 Statistical Analysis

The effects of small group size in the simulations are examined in terms of bias for all parameter estimates and their standard errors. Bias is assessed by examining whether the mean of the sampling distribution of estimates under each condition centers on the true value. If $\hat{\theta}$ is the sample estimate of the population parameter θ, then bias = $\frac{E(\hat{\theta})-\theta}{\theta}$. Bias exceeding 10% for any parameter is generally considered to be meaningful (Muthen & Muthen, 2002). The precision of the estimates is ascertained by examining the sampling distribution of standard errors for each parameter. The standard deviation of the parameter estimates in the simulations is an indicator of the population standard error when the number of replications is large (Muthen & Muthen, 2002). This is compared to the average of the estimated standard errors for each parameter estimate in the simulations, and bias in the standard errors is calculated in the same way as for the other parameter estimates, as described above.

3. Results

Table 1 presents the fixed effects and associated standard errors for the linear and non-linear multilevel models across the four simulation conditions. The true parameter values are given in parentheses in Table 1. For each of the 4 simulation conditions, the parameter estimates and the standard errors for both the continuous and discrete outcomes are estimated without bias. Even at the extremes of data sparseness (group size=2), bias in the fixed effect parameter estimates is trivial (less than 2 percent for the linear model and less than 6 percent for the non linear model) and there is no evidence of imprecision in the estimates (standard errors are in line with parameter values).

Table 1. Simulation Results for Linear and Nonlinear Multilevel Models:
Fixed Effects and Standard Errors

	Linear Multilevel Model				**Non-linear Multilevel Model**			
	Fixed Effects				Fixed Effects			
Group Size	$\hat{\gamma}_{00}$ (1.0)	$\hat{\gamma}_{10}$ (.3)	$\hat{\gamma}_{01}$ (.3)	$\hat{\gamma}_{11}$ (.3)	$\hat{\gamma}_{00}$ (1.0)	$\hat{\gamma}_{10}$ (.3)	$\hat{\gamma}_{01}$ (.3)	$\hat{\gamma}_{11}$ (.3)
2	.999	.302	.301	.302	1.030	.305	.320	.318
5	1.000	.303	.299	.301	1.002	.299	.302	.303
10	1.000	.301	.300	.299	1.000	.299	.303	.305
20	1.000	.300	.300	.299	1.000	.306	.301	.299
	Standard Errors				Standard Errors			
Group Size	$\hat{\gamma}_{00}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{01}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{00}$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{01}$	$\hat{\gamma}_{11}$
2	.040 (.039)	.042 (.042)	.039 (.042)	.042 (.043)	.176 (.176)	.157 (.165)	.142 (.149)	.162 (1.72)
5	.027 (.027)	.028 (.028)	.027 (.027)	.028 (.029)	.097 (.099)	.097 (.099)	.091 (.099)	.100 (.102)
10	.022 (.021)	.022 (.023)	.022 (.021)	.022 (.022)	.073 (.072)	.074 (.075)	.071 (.073)	.076 (.078)
20	.019 (.019)	.019 (.019)	.018 (.019)	.019 (.019)	.059 (.058)	.060 (.060)	.059 (.061)	.061 (.062)

Note: The true parameter values are given in parentheses

Table 2 presents the variance components and standard errors for both the linear and logistic multilevel models across the four simulation conditions. With very small group size (group size = 2), the group level variance components are over estimated in both the linear and non-linear models. In the linear model, the random intercept variance is overestimated by 14 percent with marginal group sizes, while both the intercept and slope variance in the non-linear model are overestimated by over 30 percent. The standard errors of these estimates are also over estimated in situations with marginal group sizes. When group size equals 2 the standard errors are biased upwards by as much as 80 percent for the group level intercept variance in the linear model, and by as much as 32 percent for the group level slope variance in the non-linear model. As a result, the power to detect significant between group variance in these models falls below .3.

Table 2. Simulation Results for Linear and Nonlinear Multilevel Models:
Random Effects and Standard Errors

	Linear Multilevel Model			**Non-Linear Multilevel Model**	
	Random Effects			Random Effects	
Group Size	$\hat{\sigma}^2$ (.5)	$\hat{\tau}_{00}$ (.044)	$\hat{\tau}_{11}$ (.044)	$\hat{\tau}_{00}$ (.366)	$\hat{\tau}_{11}$ (.366)
2	.487	.050	.048	.485	.520
5	.501	.042	.042	.358	.370
10	.500	.043	.043	.358	.366
20	.500	.043	.043	.361	.359
	Standard Errors			Standard Errors	
Group Size	$\hat{\sigma}^2$	$\hat{\tau}_{00}$	$\hat{\tau}_{11}$	$\hat{\tau}_{00}$	$\hat{\tau}_{11}$
2	.057 (.053)	.058 (.033)	.031 (.029)	.556 (.520)	.855 (.649)
5	.027 (.028)	.016 (.016)	.015 (.016)	.181 (.184)	.213 (.218)
10	.018 (.017)	.010 (.009)	.010 (.010)	.105 (.106)	.122 (.122)
20	.012 (.012)	.007 (.007)	.007 (.007)	.070 (.071)	.078 (.076)

Note: The true parameter values are given in parentheses

Table 3 presents the results for the parameter estimates following a cluster analysis of the groups. For simplicity, we only present the results for the condition where the original group size was 2, since this condition generated the only bias in the parameter estimates. After the cluster analysis, average group size increased to 6.3 observations per group, with a final total of 63,745 groups. This increase in group size generated no bias in the fixed parameter estimates or their standard errors for the continuous outcome models. However, the fixed effects in the logistic model were underestimated when using the clustered data, particularly for the group level effects (estimates biased downwards by 12% for the intercept and by 20% for γ_{01}). The standard errors of these fixed effect estimates were also consistently too short in the non linear model (biased downwards by 26-33%). Moreover, for both model types the clustering resulted in a substantial underestimation of the group level random effects and their standard errors. Both the random intercept and slope variance were biased downwards by 72-88%. For the nonlinear model, a considerable positive covariance was also introduced in the data. This underestimation of the group level variance components was accompanied by an overestimation of the within group variance in the linear models (14% positive bias), suggesting that the clustering procedure introduced artificial within group heterogeneity in the data by joining together observations from different groups.

Table 3. Simulation Results for Linear and Nonlinear Multilevel Models: Following Cluster Analysis of Group Size=2

	Linear Multilevel Model		**Non-linear Multilevel Model**	
Fixed Effects	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
$\hat{\gamma}_{00}$	1.008 (1.000)	.039 (.039)	.877 (1.000)	.121 (.176)
$\hat{\gamma}_{10}$	.279 (.300)	.041 (.042)	.301 (.300)	.122 (.165)
$\hat{\gamma}_{01}$	.301 (.300)	.039 (.042)	.240 (.300)	.108 (.149)
$\hat{\gamma}_{11}$	.291 (.300)	.040 (.043)	.267 (.300)	.115 (.172)
	Linear Multilevel Model		**Non-linear Multilevel Model**	
Random Effects	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
$\hat{\sigma}^2$	.569 (.500)	.043 (.053)	---	---
$\hat{\tau}_{00}$	.005 (.044)	.018 (.033)	.095 (.366)	.103 (.520)
$\hat{\tau}_{11}$	.012 (.044)	.020 (.030)	.096 (.366)	.101 (.649)

Note: The true parameter values are given in parentheses

4. Discussion

While concerns about data sparseness abound, very little research has been done to explicitly test the general threshold at which data sparseness becomes problematic for unbiased and efficient estimates in multilevel models. At the same time, researchers appear to be unconcerned about using cluster analysis strategies to artificially manipulate the contextual groups in the data. Using Monte Carlo simulations, this paper aimed to empirically examine the effects of data sparseness in multilevel models and to understand the effects of cluster analysis strategies designed to address data sparseness on the accuracy of model parameters.

Consistent with existing research in this area (Mok 1995; Afshartous 1995; Maas and Hox 2002; Maas and Hox 2004), multilevel models generate unbiased estimates of the fixed effects and their standard errors, even at the extremes of data sparseness. This holds for both continuous and discrete outcomes. However, there was evidence that the group level random effects and their associated standard errors are overestimated with marginal group sizes. As a result, the decreased precision in these estimates reduces the power to detect significant between group variance when group sizes are very small. When the level of clustering increases to at least 5 observations per group, there was no evidence of bias in the fixed or random effects and their standard errors.

Cluster analysis is an effective strategy for increasing group size and correcting the overestimation of the random effects in linear multilevel models. But caution should be exercised in the use of such strategies in order to prevent any underestimation of the random variance as a result of the introduction of artificial within group heterogeneity. Discrete outcomes estimated with non linear multilevel models are not robust candidates for clustering procedures. These models are extremely vulnerable to underestimation of both the fixed and random components and their standard errors. Clustering sparse data when examining discrete outcomes would lead to incorrect results and invalid conclusions about the effects of social contexts on health.

Acknowledgements

This research was supported by the Canadian Institutes of Health Research Strategic Initiative: Population and Public Health Research Methods and Tools – Pilot Project Grants.

References

Afshartous, David. (1995). "Determination of Sample Size for Multilevel Model Design". Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (Available at http://www.bus.miami.edu/~dafshart/)

Ahern, Jennifer, Pickett Kate E, Selvin, S., Abrams, B. (2003). "Preterm Birth among African American and White Women: A Multilevel Analysis of Socioeconomic Characteristics and Cigarette Smoking." *Journal of Epidemiology and Community Health*, 57:606-611.

Anderberg, Michael R. (1973). *Cluster Analysis for Applications.* New York: Academic Press, Inc.

Barr RG, Diez Roux AV, Knirsch, CA, Pablos-Mendez A. (2001). "Neighborhood poverty and the resurgence of tuberculosis in New York City, 1984-1992". *American Journal of Public Health* 19(9):1487-1493.

Bearman Peter S, Jones Jo, and Udry J. Richard. (2002). "The National Longitudinal Study of Adolescent Health: Research Design." University of North Carolina Population Center, URL: http//www.cpc.unc.edu/addhealth,.

Beland Francois, Birch Stephen, Stoddart Greg. (2002). "Unemployment and Health: Contextual-Level Influences on the Production of Health in Populations." *Social Science and Medicine*, 55:2033-2052.

Bryk Anthony S. and Raudenbush Stephen W. (1992). *Hierarchical Linear Models: Applications And Data Analysis Methods*. Newbury Park, CA: Sage.

Buka Stephen L, Brennan Robert T, Rich-Edwards Janet W, Raudenbush Stephen W, Earls Felton. (2003). "Neighborhood Support and the Birth Weight of Urban Infants." *American Journal of Epidemiology*, 157:1-8.

Cohen, Jacob. (1988). *Statistical Power Analysis*. Hillsdale, New Jersey: Erlbaum.

Clarke, Philippa and George, Linda K. (2005). "The role of the built environment in the Disablement Process." *American Journal of Public Health*; 95(11):1933-1939.

Clarke, Philippa, and Wheaton, Blair. 2007. "Addressing data sparseness in contextual population research: Using cluster analysis to create synthetic neighborhoods." *Sociological Methods and Research*, 35(3), In Press.

Cutrona, Carolyn E, Russell Daniel W, Hessling Robert M, Brown P Adama, Murry Velma. (2000). "Direct and Moderating Effects of Community Context on the Psychological Well-Being of African American Women." *Journal of Personality and Social Psychology*, 79(6):1088-1101.

Goldstein Harvey. (1995). *Multilevel Statistical Models*. London: Edward Arnold, New York: Halsted.

Gulliford, Martin C., Ukoumunne, Obioha C., Chinn Susan. (1999). "Components of Variance and Intraclass Correlations for the Design of Community-based Surveys and Intervention Studies." *American Journal of Epidemiology*, 149:876-883.

Hou Feng, Chen Jaijian. (2003). “Neighborhood Low Income, Income Inequality and Health in Toronto.” *Health Reports*, 14(2):21-34.

Kreft, Ita G. (1996). “Are Multilevel Techniques Necessary? An Overview, Including Simulation Studies”. Available online: http://www.calstatela.edu/faculty/ikreft/quarterly/quarterly.html

Kreft, Ita G. and Jan de Leeuw. (1998). *Introducing Multilevel Modeling*. Thousand Oaks: Sage Publications.

Maas, Cora JM, Hox Joop J. (2002). “Sample Sizes for Multilevel Modeling.” Pp. 1-19, in: *Social Science Methodology in the New Millenium. Proceedings of the Fifth International Conference on Logic and Methodology.* Second Expanded Edition, edited by J Blasius, J Hox, E deLeeuw, P. Schmidt. Opladen, R.G.: Leske & Budrich Verlag.

Maas, Cora JM, Hox, Joop J. (2004). “The Influence of Violence of Assumptions on Multilevel Parameter Estimates and their Standard Errors.” *Computational Statistics and Data Analysis*, 46:427-440.

Merlo, Juan, Lynch, John W., Yang, Min, Lindstrom, Martin, Ostergren Per-Olof, Rasmusen Niels K., Rastam, Lennart. (2003). “Effect of Neighborhood Social Participation on Individual Use of Hormone Replacement Therapy and Antihypertensive Medication: A Multilevel Analysis.” *American Journal of Epidemiology*, 157:774-783.

Mok, Miranda. (1995). “Sample Size Requirements for 2-Level Designs in Educational Research.” *Multilevel Modelling Newsletter* 7(2):11-15.

Muthen, Bengt O. and Muthen, Linda K. (1998). *Mplus User’s Guide*. Los Angeles, CA: Muthen and Muthen.

Muthen Linda K., and Muthen, Bengt O. (2002). “How to use a Monte Carlo study to decide on sample size and determine power”. *Structural Equation Modeling*, 4:599-620.

O’Campo P, Zue X, Wang M-C, et al. (1997). “Neighborhood risk factors for low birthweight in Baltimore: a multilevel analysis.” *American Journal of Public Health,* 87:1113-1118

Pickett Kate E, and Pearl M. (2001). “Multilevel Analyses of Neighborhood Socio-economic Context and Health Outcomes: A Critical Review.” *Journal of Epidemiology and Community Health,* 55:111-122.

Raudenbush, Stephen W., and Anthony S. Bryk. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Second Edition. Thousand Oaks: Sage Publications.

Raudenbush, Stephen W., and Robert J. Sampson. (1999). “Ecometrics: Toward a Science of Assessing Ecological Settings, With Application to The Systematic Social Observation of Neighborhoods.” *Sociological Methodology*, 29:1-41.

Robert, Stephanie A. (1998). “Community-level Socioeconomic Status Effects on Adult Health.” *Journal of Health and Social Behavior* 39(March):18-37.

Ross, Catherine E. and John Mirowsky. (2001). “Neighborhood Disadvantage, Disorder, and Health.” *Journal of Health and Social Behavior*, 42(September):258-276.

Schieman, Scott, Pearlin, Leonard I, Meersman, Stephen C. (2006). “Neighborhood disadvantage and anger among older adults: Social comparisons as effect modifiers”. *Journal of Health and Social Behavior*, 47:156-172.

Snijders, Tom A.B., & Bosker, Roel. (1999). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage Publications.

South, Scott J. and Eric P. Baumer. (2000). “Deciphering Community and Race Effects on Adolescent Premarital Childbearing.” *Social Forces*, 78(4):1379-1408.

Wheaton, Blair and Philippa Clarke. (2003). “Space meets time: Integrating temporal and contextual influences on mental health in early adulthood.” *American Sociological Review*;68:680-706.

Yen IH, Kaplan GA. (1999). “Neighborhood social environment and risk of death: multilevel evidence from the Alameda County Study.” *American Journal of Epidemiology*, 149:898-907

Yen IH, Syme SL. (1999). “The social environment and health: A discussion of the epidemiologic literature.” *Annual Review of Public Health*; 20:287-308.