



ELSEVIER

Contents lists available at ScienceDirect

## SSM - Population Health

journal homepage: [www.elsevier.com/locate/ssmph](http://www.elsevier.com/locate/ssmph)

## Article

# General and specific contextual effects in multilevel regression analyses and their paradoxical relationship: A conceptual tutorial

Juan Merlo<sup>a,b,\*</sup>, Philippe Wagner<sup>a,c</sup>, Peter C. Austin<sup>d</sup>, SV Subramanian<sup>e</sup>, George Leckie<sup>a,f</sup>

<sup>a</sup> Unit for Social Epidemiology, Department of Clinical Sciences, Faculty of Medicine, Lund University, CRC, Jan Waldenströms Street 35, SE- 214 21 Malmö, Sweden

<sup>b</sup> Center for Primary Health Care Research, Region Skåne, Malmö, Sweden

<sup>c</sup> Centre for Clinical Research Västmanland, Uppsala University, Västerås, Sweden

<sup>d</sup> Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada

<sup>e</sup> Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA

<sup>f</sup> Centre for Multilevel Modelling, University of Bristol, UK

## A B S T R A C T

To be relevant for public health, a context (e.g., neighborhood, school, hospital) should influence or affect the health status of the individuals included in it. The greater the influence of the shared context, the higher the correlation of subject outcomes within that context is likely to be. This intra-context or intra-class correlation is of substantive interest and has been used to quantify the magnitude of the *general contextual effect* (GCE). Furthermore, ignoring the intra-class correlation in a regression analysis results in spuriously narrow 95% confidence intervals around the estimated regression coefficients of the specific contextual variables entered as covariates and, thereby, overestimates the precision of the estimated *specific contextual effects* (SCEs).

Multilevel regression analysis is an appropriate methodology for investigating both GCEs and SCEs. However, frequently researchers only report SCEs and disregard the study of the GCE, unaware that small GCEs lead to more precise estimates of SCEs so, paradoxically, the less relevant the context is, the easier it is to detect (and publish) small but “statistically significant” SCEs. We describe this paradoxical situation and encourage researchers performing multilevel regression analysis to consider simultaneously both the GCE and SCEs when interpreting contextual influences on individual health.

To be relevant for public health, a context (e.g., neighborhood, school, hospital) should influence the health status of the individuals included in it. The greater this *general contextual effect* (GCE), the higher the individual correlation in health outcomes within that context should be (Merlo, Chaix, Yang, Lynch, & Rastam, 2005). This intra-context or intra-class correlation (ICC) is of interest in epidemiology for both statistical and substantive reasons. Statistically, ignoring the correlation of the outcomes of the individuals in the same context in a regression analysis results in spuriously precise standard errors around the estimated regression coefficients (Snijders & Bosker, 2012). Substantively, the ICC quantifies the size of the GCE, so the more relevant the context is for understanding individual differences in the outcome, the higher the ICC.

The existence of the ICC is the fundamental reason for applying multilevel regression models in contextual analyses. However, in practice, many analysts use multilevel regression analyses only to inform on the *specific contextual effects* (SCE). That is, they primarily focus

on quantifying cross-level associations between specific contextual variables (e.g., neighborhood socioeconomic deprivation) and the individual outcome under investigation (e.g., individual blood pressure) but disregard the information provided by the ICC, in spite of this being the very reason for applying multilevel regression analysis. This situation was described in 2007 by a scoping study of research of area effects on health reviewing publications between July 1998 and December 2005 (Riva, Gauvin, & Barnett, 2007). Our experience in the field of multilevel analysis is that reporting of the GCE has increased somewhat since then, but is still often missing in many publications. Observe that in the present study, we use the term “association” as it is traditional in Epidemiology. However, a more appropriate term could be relationship since “association” conveys an undirected relation (i.e., correlation and not regression).

We argue that when investigating contextual influences on individual outcomes, it is necessary to consider simultaneously both the GCE and SCEs. Failure to do so may lead to incorrect substantive

\* Corresponding author at: Unit of Social Epidemiology, Department of Clinical Sciences, Faculty of Medicine, Lund University, CRC, Jan Waldenströms Street 35, SE- 214 21 Malmö, Sweden.

E-mail address: [juan.merlo@med.lu.se](mailto:juan.merlo@med.lu.se) (J. Merlo).

<https://doi.org/10.1016/j.ssmph.2018.05.006>

Received 20 March 2018; Received in revised form 12 May 2018; Accepted 15 May 2018

2352-8273/© 2018 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

conclusions because of the presence of a paradoxical relationship between the GCE and SCEs. Namely, the existence of a small crude GCE (i.e., a small ICC before the inclusion of cluster level variables in the multilevel model) suggests that the context under investigation is of little relevance for the individual outcome but, simultaneously, this small GCE allows any SCEs to be estimated relatively precisely. In other words, all else being equal, the 95% confidence interval of an estimated SCE narrows (and the statistical “significance” increases) as the GCE decreases. Thus, paradoxically, the less relevant the general context is for the individual outcome (i.e., the lower the GCE is), the more likely researchers may be to report trivially small SCE as “statistically significant”. Therefore, researchers who exclusively focus on the analysis of SCE may arrive at the paradoxical conclusion that the general context is relevant when in fact it is not.

Similarly, the existence of an initially large crude GCE (i.e., a large ICC before the inclusion of cluster level variables in the multilevel model) suggests that the context under investigation seems of high relevance for the individual outcome. However, simultaneously, this large ICC may lead to imprecisely estimated SCEs if the GCE remains large after the inclusion of the cluster level variable in the model (i.e., the cluster level variable does not considerably explain the cluster variance). Therefore, the paradox here is that the more relevant the general context is for the individual outcome, the easier it is to identify inconclusive (i.e., “statistically non-significant”) associations between specific contextual variables and the individual outcome. Consequently, as in the previous case, the exclusive study of SCEs in this scenario may lead to the paradoxical conclusion that the context is irrelevant when in fact it is not. This paradoxical relationship between the GCE and SCEs is not the expression of a statistical bias but rather an argument for interpreting simultaneously SCEs together with the crude and adjusted GCEs.

In this study, we aim to provide appropriate background information for the epidemiological reader to understand the paradoxical relationship between the GCE and SCEs. We start our article with a detailed, but non-mathematical explanation of the difference between *general* and *specific* contextual effects in multilevel analyses. Then, in order to understand the paradox, we introduce the concepts of the *design effect* and *effective sample size* and provide several illustrative examples. As this is a methodological paper, in our examples we assume that individuals are randomly assigned to clusters (no confounding), which is a strong assumption in the case of observational studies of neighbourhoods (Oakes, 2004). We also assume that the reader is aware of the problems of estimating causal effects from epidemiological data (Hernan & Robins, 2006).

Our study ends with the recommendation of simultaneously interpreting GCEs and SCEs when investigating the influence of the context on individual health outcomes. In the remainder of the paper, we will often refer to neighborhoods as our example of clusters or contexts, but our ideas are fully applicable to other context definitions such as hospitals or schools.

## 1. General and specific contextual (observational) effects in multilevel analyses

Multilevel regression analyses (i.e., mixed-effects models, random effects models, hierarchical regression models) are now widely applied to the analysis of longitudinal and cross-sectional data. In longitudinal data settings, repeated measurements (particularly of physiological characteristics such as blood pressure) are highly correlated within individuals, but this correlation is not normally of substantive interest; it is simply a nuisance parameter that needs to be taken into account in order to obtain correct standard errors. However, when multilevel regression techniques are applied in cross-sectional settings to investigate contextual or cluster (e.g., the neighborhoods’, schools’, hospitals’) effects on individual health, the magnitude of the ICC is substantively interesting.

From a repeated measures perspective, the individuals could be considered as repeated measurements of the cluster. The problem is that while individuals are very well-defined socio-biological systems and the ICC for repeated measurements within individuals is in most occasions high, many clusters (e.g., neighborhoods) are not well-defined systems and the ICC for individual outcomes is consequently low. Neighborhoods are frequently delimited by administrative geographical boundaries that do not necessarily capture the relevant physical or sociological contexts that influences individual health (Merlo, Ohlsson, Lynch, Chaix, & Subramanian, 2009). The point is that knowledge of the size of the individual correlation within neighborhoods (i.e., the ICC) is no longer a statistical nuisance but bears important substantive information on the validity of the neighborhood definition as a valid contextual construct that influences the individual outcome. As expressed by Rodriguez and Goldman in 1995 (Rodriguez & Goldman, 1995) (page 74):

“Estimates of the extent to which observations within a given group are correlated with one another are valuable not only for obtaining improved estimates of fixed effects and their standard errors but also for yielding important substantive information. In particular, estimates of the extent of similarity of observations within a cluster, with and without the introduction of a set of control variables, may provide insights into the influence of group level effects on individual behavior and the pathways through which these effects operate”.

If the neighborhood context has a strong influence for a specific health outcome, the individual correlation of outcomes within neighborhoods will be high: the individuals sharing the same neighborhood context will display a strong similarity for that specific health outcome. On the other hand, if the individual correlation within neighborhoods is very low or absent, the neighborhoods will be more akin to random samples from the overall population of individuals than meaningful contexts influencing the specific individual outcome under investigation. We refer to this general influence of the context as the GCE and we quantify it using measures of variance and clustering. In contrast, we estimate SCEs via cross-level measures of average association, for instance, linear regression coefficients or in the case of multilevel logistic regression, odds ratio (OR) between specific neighborhood characteristics and the individual outcome. The SCEs also explain the neighborhood variance and, thereby, shed light on the mechanisms mediating the GCE under investigation (Austin & Merlo, 2017; Merlo, 2003; Merlo et al., 2006).

During the last two decades, different terms have been used to describe this GCE idea. As indicated above, Rodriguez and Goldman describe the GCE as “group level effects” (Rodriguez & Goldman, 1995). Boyle investigated “place effects” (Boyle & Willms, 1999), while Petronis & Anthony identified the existence of “a different kind of contextual effect” (Petronis & Anthony, 2003), Merlo quantified “population effects” (Merlo, Asplund, Lynch, Rastam, & Dobson, 2004) and distinguished between measures of variance and measures of association linking the statistical concept of clustering to the idea of “contextual phenomenon” (Merlo, 2003; Merlo et al., 2005), and Subramanian described GCEs as “common ecological effects” (Subramanian, 2004; Subramanian, Glymour, & Kawachi, 2007).

In summary, SCEs are measured by estimating the cross-level association between specific neighborhood characteristics (e.g., neighborhood socioeconomic deprivation) and individual outcomes (e.g., systolic blood pressure). In contrast, GCEs are quantified by measures of clustering such as the ICC (Merlo et al., 2006). GCEs estimate the effects of neighborhood contexts on individual outcomes, without reference to any of specific neighborhood characteristics other than the very boundaries defining neighborhoods. Focusing on the ICC as a measure of the GCE, when outcomes are continuous the ICC can be formulated as,

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \tag{1}$$

where  $\sigma_u^2$  is the variance in individual outcomes which lies between neighborhoods and  $\sigma_e^2$  the variance between individual outcomes within neighborhoods. The ICC therefore quantifies the share of the total individual outcome variance ( $\sigma_u^2 + \sigma_e^2$ ) that is due the contextual level ( $\sigma_u^2$ ). The ICC needs to be taken into account when interpreting the SCEs for both statistical and substantive reasons. Statistically, failure to consider the ICC results in spuriously precise regression coefficients. Substantively, the ICC measures the GCE.

The focus of our paper is on the continuous case and we need to note that even if the categorical case is similar in principle, it also raises other issues that are not considered in the present paper. For binary outcome variables, the calculation of the intra-neighborhood correlation is less straightforward than for continuous outcomes. A range of approaches have been proposed (Austin & Merlo, 2017; Browne, Subramanian, Jones, & Goldstein, 2005; Goldstein, Browne, & Rasbash, 2002; Li, Gray, & Bates, 2008; Merlo et al., 2006), the most popular of which appears to be that based on the latent response formulation of the multilevel logistic regression model. Using this approach, the resulting ICCs for binary outcomes are interpreted in the same way as in the continuous outcome case. In addition, the GCE can be assessed by other measures of clustering including the pairwise odds ratio (PWOR) (Petronis & Anthony, 2003) as well as by measures of discriminatory accuracy (Merlo, Wagner, Ghith, & Leckie, 2016; Wagner & Merlo, 2013, 2014). The GCE can also be assessed by measures of heterogeneity like the median odds ratio (MOR) (Larsen & Merlo, 2005; Larsen, Petersen, Budtz-Jorgensen, & Endahl, 2000), the median hazard ratio (MHR) (Austin, Wagner, & Merlo, 2017b) and the median (incidence) rate ratio (MRR) (Austin, Stryhn, Leckie, & Merlo, 2017a). However, the reader needs to be aware the quantification of the components of variance and the GCE for discrete outcomes is possible but more complex than in the case of continuous outcomes as analyzed in the present tutorial.

## 2. The paradoxical relationship between general and specific contextual effects in multilevel regression analysis

Having clarified the conceptual distinction between the GCE and SCEs, we proceed to explain how the exclusive focus on SCEs without the simultaneous consideration of GCEs may lead to misinterpretations of the contextual influence of clusters on individual outcomes. We first introduce the concepts of the *design effect* and *effective sample size* and then we provide several illustrative examples.

### 2.1. On the design effect and the effective sample size

To help the reader's intuition on why the ICC in multilevel data reduces the *effective sample size* and thereby the statistical certainty of the estimated regression coefficients, imagine that we are studying individual height in a sample of 50 pairs of monozygotic (identical) twins. In this most extreme of examples, it is very intuitive that we must have an effective sample size of around 50 independent measurements, rather than 100 because the intra-twin pair correlation is close to 0.9 (90% of the individual variation lies between twin pairs, only 10% within twin pairs). That is, statistically, every twin pair contributes approximately only as much independent information as one individual. To quantify this phenomenon we first calculate the design effect (Gulliford, Ukoumunne, & Chinn, 1999) as

$$\text{Design effect} = 1 + (n - 1) \times ICC$$

where  $n$  is average cluster size.

In our twins example

$$\text{Design effect} = 1 + (2 - 1) \times 0.9$$

$$\text{Design effect} = 1.9$$

Thereafter, using the design effect we can calculate the “effective sample size” as

$$\text{Effective sample size} = (J \times n / \text{Design effect})$$

where  $J$  is the number of clusters.

In our twins' example.

$$\text{Effective sample size} = (100 / 1.9)$$

$$\text{Effective sample size} = 53$$

Thus, our sample of 100 twins provides as much information on height as does a sample of 53 unrelated individuals.

The twin example is an extreme but illustrative case as the high ICC has clear biological causes. However, there are many less extreme but analogous situations in epidemiological practice (e.g., blood lipid levels within family households, as families share a similar diet and genetic traits). Indeed, the terms “design effect” and “effective sample size” are customarily used by statisticians in epidemiology and other fields performing cluster randomized trials and complex population surveys (Cornfield, 1978; Gulliford et al., 1999).

We will apply these concepts in the next section to understand the paradoxical relationship between GCEs and SCEs in multilevel regression analysis.

### 2.2. An illustrative example with three scenarios

In multilevel regression analysis of individuals nested within neighborhoods, we often estimate both associations between individual level characteristics and individual outcomes and cross-level associations between neighborhoods characteristics and individual outcomes. These cross-level associations represent the SCEs described in the previous section. The existence of a GCE decreases the effective sample size and therefore reduces the precision of the estimated individual and, especially, cross-level associations (Snijders & Bosker, 2012). This situation is well-known. However, what is less well known is that this phenomenon is also responsible for an interpretational paradox between the GCE and SCEs.

Below, we illustrate this paradox using a fictitious example analyzing neighborhoods' effects on systolic blood pressure in 1250 individuals residing in 50 neighborhoods with 25 residents in each neighborhood. We investigate neighborhoods since, according to our theories, we assume that neighborhoods exert general influences that conditions individual systolic blood pressure level. Therefore, we anticipate the existence of a GCE (i.e., an intra-neighborhood correlation of individual systolic blood pressure). In addition, our hypothesis is that the neighborhood GCE is mediated by the socioeconomic characteristics of the neighborhoods and so we expect that a neighborhood socioeconomic deprivation variable will explain part of the between-neighborhood variance and be associated with individual systolic blood pressure. In summary, we aim to quantify the crude and adjusted neighborhood GCEs as well as the cross-level association between neighborhood socioeconomic deprivation and individual systolic blood pressure level (i.e., the SCE).

Given the above scenario, we should conduct a multilevel regression analysis that takes into account the intra-neighborhood correlation. In contrast, had we naively applied a conventional linear regression analysis at the individual level, the analysis would implicitly assume that the sample consisted of 1250 independent individuals. Consequently, by ignoring the presence of an ICC, we would overestimate the precision of the estimated effects, and obtain spuriously narrow 95% confidence intervals around the regression coefficients of the associations.

To illustrate the results of this fictitious study we describe three different scenarios A, B and C. For simplicity, the outcome (individual systolic blood pressure) and covariate (neighborhood socioeconomic deprivation) have been standardized to have mean 0 and variance 1 in

**Table 1**

Numerical representation of three different Scenarios (A, B and C) obtained from a Ready Reckoner (Available In The Appendix With Extended Information). The fictitious example analyzes neighborhoods’ observational effects on systolic blood pressure in a sample of 1250 individuals residing in 50 neighborhoods with 25 residents in each. The example represents a multilevel linear regression analysis. Model 1 is a two-level variance-components model. Model 2 extends Model 1 by entering the cluster-level variable neighborhood socioeconomic deprivation. The outcome and covariate are standard normalized.

	Scenario A		Scenario B		Scenario C	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
<b>Specific contextual effect (SCE)</b>						
Neighborhood socioeconomic deprivation						
Regression coefficient (1 SD increase)		0.671		0.097		0.134
95% Confidence interval		0.629 0.712		0.042 0.153		-0.053 0.320
p-value of slope coefficient		0.000		0.000		0.161
<b>General contextual effect (GCE)</b>						
Intraclass correlation coefficient (as a %)	45.00%	0.05%	1.00%	0.05%	45.00%	44.00%
Neighborhood-level variance	0.450	0.0003	0.010	0.0005	0.450	0.432
Individual level variance	0.550	0.550	0.990	0.990	0.550	0.550
Total residual variance	1.000	0.550	1.000	0.990	1.000	0.990
Explained neighborhood variance		99.9%		95.0%		4.0%
Effective sample size	106	1235	1008	1235	106	108
Statistical power		1.000		0.931		0.288

each scenario. All statistics presented for these scenarios have been calculated using a “ready reckoner” or interactive excel workbook, which allows readers to interactively explore the relationship between the GCE and SCEs in a Microsoft Excel spreadsheet (Supplemental digital content 1). We describe the model equations and formula for all derived statistics, which appear in the ready reckoner in detail in the Appendix.

To illustrate the results of this fictitious study we describe three different scenarios A, B and C. All statistics presented for these scenarios have been calculated using a “ready reckoner” or interactive excel workbook, which allows readers to interactively explore the relationship between the GCE and SCEs in a Microsoft Excel spreadsheet (Supplemental digital content 1). We describe the model equations and formula for all derived statistics, which appear in the ready reckoner in detail in the Appendix. We present a summary of the results in Table 1.

**2.2.1. Scenario A**

In this scenario, the initial (i.e., crude or unadjusted) ICC in model 1 is equal to 45%. This high GCE suggests that the context as a whole is very relevant for understanding individual differences in systolic blood pressure. Subsequently, we include the neighborhood socioeconomic deprivation variable in model 2, and it explain almost all (i.e., 99.9%) of the neighborhood variance, so the conditional ICC decreases considerably and becomes only 0.05%. That is, the GCE is almost completely explained by the SCE of neighborhood deprivation. The effective sample size of individuals increases from 106 in model 1 to 1235 in model 2. This effective sample size is close the actual number of 1250 individuals and it goes hand-in-hand with the very low adjusted or conditional ICC. As expected, the neighborhood socioeconomic deprivation variable appears strongly associated with individual hypertension (slope coefficient = 0.671) and the 95% confidence interval: 0.629, 0.712 is narrow. Obviously, the SCE is statistically highly significant.

Scenario A represents an ideal situation where the neighborhood context via socioeconomic deprivation strongly conditions individual systolic blood pressure.

**2.2.2. Scenario B**

Imagine a second situation in which the initial ICC is only 1% in model 1, which is a very low GCE and suggests that the context is rather unimportant. Because of the small initial ICC, the effective sample size is very large (1008). Thereafter we include the neighborhood variable in model 2 and it explains a large proportion (i.e., 95%) of the neighborhood variance so that the adjusted or conditional ICC becomes

0.05%. Furthermore, the neighborhood socioeconomic deprivation variable is conclusively associated with individual systolic blood pressure. Even if this SCE is not as large as in scenario A (slope coefficient = 0.097), the 95% confidence interval: 0.042, 0.153 is narrow and the effective sample size of 1235 is as large as in the scenario A.

Observe, however, that the initial GCE in model 1 was very low and that even if the neighborhood socioeconomic deprivation variable explains a large share of the neighborhood variance, this variance was very small already in model 1.

**2.2.3. Scenario C**

Now imagine a third situation similar to scenario A with an initially very large ICC (i.e., 45%) in model 1. This high GCE means that once again the neighborhood context strongly conditions individual differences in systolic blood pressure. However, in model 2, the neighborhood socioeconomic deprivation variable only explains 4% of the initial neighborhood variance so the ICC remains very high (i.e., 44%). In this case, the effective sample size in model 2 is 108, which is little changed from the effective sample size of 106 in model 1. We see that even if the slope of the neighborhood variable (i.e., the SCE) is rather high (slope coefficient = 0.134), the 95% confidence interval: -0.053, 0.320 is wide and inconclusive and the association is not statistically significant because the effective sample size is small.

An investigator that only focuses on SCEs will conclude that the neighborhood context is relevant in both the scenario A and scenario B but not in scenario C. However, applying the concepts highlighted in this study, we conclude that the neighborhood context is relevant in scenarios A and C but not in scenario B. Furthermore, we conclude that even though the neighborhood socioeconomic deprivation variable is statistically significant in scenario B, the association is rather irrelevant given the tiny initial GCE.

The interesting reader may find empirical examples of the scenarios A, B and C elsewhere (Larsen & Merlo, 2005; Merlo et al., 2004; Merlo et al., 2001; Merlo et al., 2016). Without performing a systematic literature review, our experience suggests that many of the published studies on neighborhood and health are similar to scenario B, in which relatively small SCEs are found to be statistically significant and presented as evidence that the neighborhood context more generally is relevant when in fact the initial GCE is very weak.

**3. Discussion and conclusions**

We illustrate how an exclusive focus on SCEs may lead to misleading conclusions as to the general relevance of the context for the individual

outcome. It is rather common that a neighborhood level variable that is weakly associated with the outcome may show a “statistically significant” SCE only when the initial GCE is weak. In other words, the less relevant the neighborhood context for the individual outcome, the easier it is to find a “statistically significant” association between a specific neighborhood characteristic and the individual outcome and – erroneously – conclude that (based on the SCE) the neighborhood context is relevant. This paradox may lead research to focus on the wrong context as it is easier to obtain tiny but “statistically significant” associations (and get those findings published) when the context is less relevant or incorrectly specified. Alternatively, we could find a theoretically pertinent neighborhood level variable that is statistically not “significantly” associated with the outcome under investigation and conclude that the neighborhood context more generally is not relevant. However, the real reason for the lack of SCE could be that there is a substantial remaining unexplained GCE. This unexplained GCE informs the researcher that the context is relevant but that the pertinent neighborhood-level covariates are yet to be identified. That is, we need to consider the size of the initial GCE before including cluster specific variables to quantify SCEs.

### Grants and/or financial support

This work was mainly supported by the Swedish Research Council (VR #2013–2484, PI: Merlo) and by a Career Investigator award from the Heart and Stroke Foundation (Dr. Peter Austin).

### Conflict of interest statement

Not declared.

### Financial disclosure statement

Academic work supported by public funds.

### Ethics

This study does not require ethics approval since is based on simulated data.

### Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.ssmph.2018.05.006>.

### References

- Austin, P. C., & Merlo, J. (2017). Intermediate and advanced topics in multilevel logistic regression analysis. *Stat Med*, *36*, 3257–3277.
- Austin, P. C., Stryhn, H., Leckie, G., & Merlo, J. (2017a). Measures of clustering and heterogeneity in multilevel Poisson regression analyses of rates/count data. *Statistics in Medicine*.
- Austin, P. C., Wagner, P., & Merlo, J. (2017b). The median hazard ratio: A useful measure of variance and general contextual effects in multilevel survival analysis. *Statistics in Medicine*, *36*, 928–938.
- Boyle, M. H., & Willms, J. D. (1999). Place effects for areas defined by administrative boundaries. *American Journal of Epidemiology*, *149*, 577–585.
- Browne, W. J., Subramanian, S. V., Jones, K., & Goldstein, H. (2005). Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*, 599–613.
- Cornfield, J. (1978). Randomization by group: A formal analysis. *American Journal of Epidemiology*, *108*, 100–102.
- Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in generalised linear multilevel models. *Understanding Statistics*, *1*, 223–232.
- Gulliford, M. C., Ukoumunne, O. C., & Chinn, S. (1999). Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: Data from the Health Survey for England 1994. *American Journal of Epidemiology*, *149*, 876–883.
- Hernan, M. A., & Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health*, *60*, 578–586.
- Larsen, K., & Merlo, J. (2005). Appropriate assessment of neighborhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression. *American Journal of Epidemiology*, *161*, 81–88.
- Larsen, K., Petersen, J. H., Budtz-Jorgensen, E., & Endahl, L. (2000). Interpreting parameters in the logistic regression model with random effects. *Biometrics*, *56*, 909–914.
- Li, J., Gray, B. R., & Bates, D. M. (2008). An empirical study of statistical properties of variance partition coefficients for multi-level logistic regression models. *Communications in Statistics – Simulation and Computation*, *37*, 2010–2026.
- Merlo, J. (2003). Multilevel analytical approaches in social epidemiology: Measures of health variation compared with traditional measures of association. *Journal of Epidemiology and Community Health*, *57*, 550–552.
- Merlo, J., Asplund, K., Lynch, J., Rastam, L., & Dobson, A. (2004). Population effects on individual systolic blood pressure: A multilevel analysis of the World Health Organization MONICA Project. *American Journal of Epidemiology*, *159*, 1168–1179.
- Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., et al. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health*, *60*, 290–297.
- Merlo, J., Chaix, B., Yang, M., Lynch, J., & Rastam, L. (2005). A brief conceptual tutorial of multilevel analysis in social epidemiology: Linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of Epidemiology and Community Health*, *59*, 443–449.
- Merlo, J., Ohlsson, H., Lynch, K. F., Chaix, B., & Subramanian, S. V. (2009). Individual and collective bodies: Using measures of variance and association in contextual epidemiology. *Journal of Epidemiology and Community Health*, *63*, 1043–1048.
- Merlo, J., Ostergren, P. O., Hagberg, O., Lindstrom, M., Lindgren, A., Melander, A., et al. (2001). Diastolic blood pressure and area of residence: Multilevel versus ecological analysis of social inequity. *Journal of Epidemiology and Community Health*, *55*, 791–798.
- Merlo, J., Wagner, P., Ghith, N., & Leckie, G. (2016). An original stepwise multilevel logistic regression analysis of discriminatory accuracy: The case of neighbourhoods and health. *Plos One*, *11*, e0153778.
- Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology. *Social Science & Medicine*, *58*, 1929–1952.
- Petronis, K. R., & Anthony, J. C. (2003). A different kind of contextual effect: Geographical clustering of cocaine incidence in the USA. *Journal of Epidemiology and Community Health*, *57*, 893–900.
- Riva, M., Gauvin, L., & Barnett, T. A. (2007). Toward the next generation of research into small area effects on health: A synthesis of multilevel investigations published since July 1998. *Journal of Epidemiology and Community Health*, *61*, 853–861.
- Rodriguez, G., & Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society A*, *158*, 73–78.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Los Angeles: Sage.
- Subramanian, S. V. (2004). The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Social Science & Medicine*, *58*, 1961–1967.
- Subramanian, S. V., Glymour, M. M., & Kawachi, I. (2007). Identifying causal ecological effect on health: A methodological assessment. In S. Galea (Ed.), *Macrosocial determinants of population health* (pp. 301). New York, NY: Springer.
- Wagner, P., & Merlo, J. (2013). Measures of discriminatory accuracy in multilevel analysis. *European Journal of Epidemiology*, *28*, 135.
- Wagner, P., & Merlo, J. (2014). Discriminatory accuracy of a random effect in multilevel logistic regression. *International Journal of Epidemiology*, *44*, i49–i50.