



LUND UNIVERSITY

Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association.

Merlo, Juan

Published in:
Journal of Epidemiology and Community Health

2003

[Link to publication](#)

Citation for published version (APA):

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(8), 550-552.
<http://jech.bmjjournals.com/cgi/content/extract/57/8/550>

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association

J Merlo

Considering both distribution and determinants of health

In this issue of the journal Jennifer Ahern *et al*¹ present the results of a multilevel analysis showing the increased likelihood of preterm birth affecting both African American and white women living in a neighbourhood with deprived socioeconomic characteristics. This increased risk was independent of individual cigarette smoking and modified by socioeconomic characteristics of the women. The authors, taking a multilevel perspective, concluded that examining both neighbourhood and individual socioeconomic factors in combination with behavioural and biological factors is the most adequate way to study determinants of preterm delivery.

MULTILEVEL ANALYSIS WITHOUT MULTILEVEL REGRESSION—IS THE INTRA-NEIGHBOURHOOD CORRELATION* A NUISANCE?

The study of Ahern *et al* contributes to the growing stream of current multilevel analysis in modern health epidemiology. However, the analytical approach of Ahern's multilevel analysis does not apply *multilevel regression* (synonymous

with *hierarchical regression*)²⁻⁵ for statistical modelling. The authors describe the association between preterm birth and neighbourhood variables by population-average regression techniques that account for intra-neighbourhood correlation using a method called generalised estimating equations (GEE).^{6,7} In this way the authors simply aim to provide acceptable estimates for the standard errors around point estimates (that is, odds ratios, 95% CI), treating the intra-neighbourhood correlation as a "nuisance" that needs to be adjusted in the analysis but not explicitly investigated.

Analogously to the study of Ahern, other social epidemiologists have adopted a similar analytical approach, applying SUDAAN statistical software (<http://www.rti.org/>) to perform multilevel analysis. As in Ahern's study, SUDAAN analyses also consider the variance structure of the data as a necessary "nuisance". These authors' analytical approach and the estimation of the association between neighbourhood characteristics and health are, however, appropriate and formally correct.

*The most fundamental reason for applying special statistical techniques in multilevel analysis is the existence of intraclass (intra-neighbourhood) correlation. The intraclass correlation is a measure of the degree of similarity among the outcomes of members of the same neighbourhood. Individuals living in the same neighbourhood may be more similar to each other than individuals living in other neighbourhoods, as they share a number of economic, social, and other neighbourhood characteristics that may condition similar health status. In this sense neighbourhoods can be considered as "clusters" of individuals sharing a common propensity for similar outcome within clusters. More technically, the intraclass correlation is a *variance partition coefficient* that indicates the proportion of the total variance (V)—that is, the sum of 1st level (individual) and 2nd level (neighbourhood) variances in a health outcome that it is accounted for by the 2nd level variance.⁸ Intraclass correlation needs to be accounted for in regression analysis, as in the study of Jennifer Ahern *et al*.¹ Otherwise the lack of independence, arising from two sources of variation at different levels (individual and neighbourhood) of the data hierarchy contradicts the assumption for performing traditional regression analysis. If ICC is not considered, the study sample is artificially "inflated" and the standard error of neighbourhood variables underestimated. One can imagine 100 neighbourhoods with some 50 individuals each (that is, population size=5000 individuals). If the individuals within each neighbourhood are exactly similar to each other, but completely different from the individuals in the other neighbourhood (intra-neighbourhood correlation=100%), the effective number of individuals would be 100 rather than 5000. In other words, if the ICC=100%, the effective population size will be the number of neighbourhoods, rather than the number of individuals.

$$ICC = \frac{V_{2nd\ level}}{V_{2nd\ level} + V_{1st\ level}}$$

Is this the end of the story? Is the only reason for applying complicated statistical techniques so that correct confidence intervals may be obtained? Is the intra-neighbourhood correlation only a "nuisance" that needs to be controlled but not investigated? Is knowledge regarding multilevel measures of health variation, like intra-neighbourhood correlation, irrelevant in social epidemiology?

MULTILEVEL MEASURES OF HEALTH VARIATION

Within social epidemiology, explicit knowledge about intra-neighbourhood correlation is of importance for substantive epidemiological reasons. Estimation of the extent to which individuals within a given neighbourhood are correlated with one another in relation to health (the concept of intra-neighbourhood correlation) yields important information by itself. The more the health of the individuals within a neighbourhood are alike (as compared with individuals in other neighbourhoods), the more likely it is that the determinants of individual health have to do directly with the contextual environment of the neighbourhood,⁴ and/or that strong social processes of contextual/geographical segregation are taking place—that is, similar types of individuals choose or are forced to reside in a given neighbourhood.

The investigation of multilevel measures of health variation (for example, slope variance, modelling of variance, variance partition coefficient, and intra-neighbourhood correlation) yield more extended and sophisticated information than traditional measures of association (for example, regression coefficients, odds ratios).^{8,9} For multilevel logistic regression Larsen has proposed using a median odds ratio (MOR) measure that reflects the second level (that is, neighbourhood) variance and can be used to quantify area effects on individual health.¹⁰ This author has also proposed the use of an interval odds ratio (IOR) that integrates neighbourhood variation in measures of association. MOR and IOR are intuitive and easy to be interpreted in terms of well known odds ratios. In general, the use of measures of health variation is a rather new but promising methodological approach that needs to be developed in social epidemiology.

When studying individuals nested within neighbourhoods, an ICC=0% suggests that the areas are not important determinant of individual health, as the neighbourhoods resemble random samples from the whole population.

INTERVENTION FOCUSED ON PLACES INSTEAD OF PEOPLE

Estimation of the extent to which individuals within a given neighbourhood are correlated with one another in relation to health (that is, the concept of intra-neighbourhood correlation) has value in the context of ideas about the efficacy of focusing intervention on places instead of people. Traditional measures of association like odds ratios can say nothing about how individual health variation in the population differs or correlates between neighbourhoods. For example, if an intervention were to focus on a given selection of "high risk" neighbourhoods, when in fact neighbourhood variation was actually a very small part of the total variation, then a very large number of high risk individuals would be missed simply because they reside in apparently middle and low risk neighbourhoods.¹¹ In other words, when the intra-class correlation is small, focusing intervention on places may be a rather inefficient strategy. Therefore, by basing our investigation on the size of the intra-neighbourhood correlation, we can evaluate the relative importance of the neighbourhood level for different kinds of outcomes, and can promote resources for community intervention for those health outcomes that are largely determined by the neighbourhood. Traditional measures of association such as odds ratios thus provide an incomplete epidemiological basis for decision making in public health interventions. Nevertheless, analysis of traditional measures of association has been the approach most commonly used in multilevel population health research.¹⁻¹²

MEASURES OF HEALTH VARIATION AND TRADITIONAL MEASURES OF ASSOCIATION GIVE COMPLEMENTARY INFORMATION

It is possible to find large traditional measures of effect (regression coefficients, odds ratios) side by side with smaller measures of health variation (neighbourhood variance and intra-neighbourhood correlation).¹³ Even more, neighbourhood variables tend to be more "significant" and have smaller confidence intervals when the intra-neighbourhood correlation is low. We need to understand that large odds ratios and a low intraclass correlation are not contradictory facts, but they give different and complementary information.¹⁴

Natural neighbourhood differences, even when very small, may give enough contrast of exposure to detect an association and this association is rather independent of the individual variation. The accompanying figure shows that it is possible to imagine a situation when an

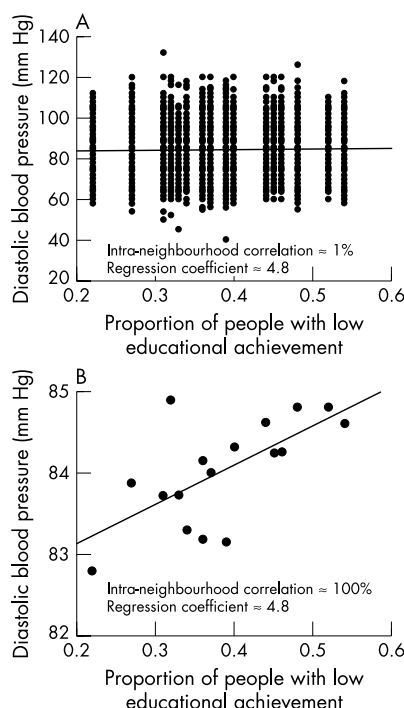


Figure 1 (A) and (B) Present two multilevel analyses showing the exact same association (regression coefficient, $\beta=4.8$) between diastolic blood pressure and proportion of people with low educational achievement. However, the size of the intra-neighbourhood correlation ranged from less than 1% (A) to 100% (B). In the first case (A) the areas do not differ more than random samples taken from the whole population, and the geographical environment has almost no effect on the individual outcome. In the second case (B), the clustering of persons in relation to blood pressure is total, and the geographical environment completely influences individual outcome. Despite the large disparity in the size of the intra-neighbourhood correlation, the size of the regression coefficients is similar (that is, $\beta=4.8$) in both cases. A similar figure has been previously published in the *Journal of Epidemiology and Community Health* and is reproduced here with permission.¹¹

evident association (regression coefficient, $\beta=4.8$) between neighbourhood proportion of people with low educational achievement and blood pressure coexists when the intraclass correlation is very large, but also when it is very small. It is obvious that we can observe the same means with very different variation around these means. In the analysis of traditional measures of association we focus on fixed mean parameters. However, in analysing components of health variation we mainly focus our attention on variance around the means.¹¹

CONCLUSIONS

Certainly broad social and economic forces generate differences among neighbourhoods that shape the distribution of health outcomes.¹⁵ Strategies of

disease prevention need to combine person centred approaches with approaches aimed at changing residential environments.¹⁶ For this task, traditional measures of association (for example, regression coefficients, odds ratios) between neighbourhood socioeconomic characteristics and individual health are a relevant approach to understanding cross level effect pathways and social determinants of health.¹⁷ However, when it comes to evaluating multilevel risk distribution and the public health relevance of specific administrative boundaries¹⁸ (for example, districts, municipalities, neighbourhoods) on different individual health outcomes, multilevel measures of health variation (for example, intra-neighbourhood correlation) present themselves as a new epidemiological approach that may prove very useful in social epidemiology.

ACKNOWLEDGEMENTS

I want to express my sincere gratitude to Bo Gullberg and Klaus Larsen for revising and commenting on the manuscript.

J Epidemiol Community Health 2003;57:550-552

Author's affiliations

J Merlo, Department of Community Medicine, Malmö University Hospital, Faculty of Medicine, Lund University, S-205 02 Malmö, Sweden

Correspondence to: Dr J Merlo; Juan.Merlo@smi.mas.lu.se

REFERENCES

- Ahern J, Pickett KE, Selvin S, et al. Preterm birth among African American and white women: a multilevel analysis of socioeconomic characteristics and cigarette smoking. *J Epidemiol Community Health* 2003;57:606-11.
- Goldstein H. *Multilevel statistical models*. 3rd edn. London: Hodder Arnold, 2003.
- Raudenbush SW, Bryk AS. *Hierarchical linear models: applications and data analysis methods*. 2nd edn. London: Sage Publications, 2002.
- Snijders TAB, Bosker RJ. *Multilevel analysis-an introduction to basic and advanced multilevel modeling*. 1st edn. Thousand Oaks, CA: Sage Publications, 1999.
- Leyland AH, Goldstein H. *Multilevel modeling of health statistics*. Chichester: Wiley, 2001.
- Hanley JA, Negassa A, Edwards MD, et al. Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am J Epidemiol* 2003;157:364-75.
- Goldstein H, Browne W, Rasbash J. Marginal vs hierarchical models. *Multilevel modelling of medical data*. 2002. (<http://www.ioe.ac.uk/hgpersonal/multilevel-medical-tutorial.pdf>)
- Rasbash J, Browne W, Goldstein H, et al. Modelling the variance as a function of explanatory variables. *A user's guide to MLwiN*. London: Multilevel Models Project, Institute of Education, University of London, 2000:77-88.
- Goldstein H, Browne W, Rasbash J. Partitioning variation in multilevel models. 2002. (<http://www.ioe.ac.uk/hgpersonal/Variance-partitioning.pdf>)

- 10 **Larsen K**, Petersen JH, Budtz-Jorgensen E, *et al.* Interpreting parameters in the logistic regression model with random effects. *Biometrics* 2000;**56**:909–14.
- 11 **Merlo J**, Ostergren PO, Hagberg O, *et al.* Diastolic blood pressure and area of residence: multilevel versus ecological analysis of social inequity. *J Epidemiol Community Health* 2001;**55**:791–8.
- 12 **Pickett KE**, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health* 2001;**55**:111–22.
- 13 **Merlo J**, Lynch JW, Yang M, *et al.* Effect of neighbourhood social participation on individual use of hormone replacement and anti-hypertensive medications: a multilevel analysis. *Am J Epidemiol* 2003;**157**:774–83.
- 14 **Sampson RJ**, Morenoff JD, Gannon-Rowley T. Assessing “neighbourhood effects”: social processes and new directions in research. *Annu Rev Sociol* 2002;**28**:443–78.
- 15 **Diez Roux AV**, Merkin SS, Arnett D, *et al.* Neighborhood of residence and incidence of coronary heart disease. *N Engl J Med* 2001;**345**:99–106.
- 16 **Macintyre S**, Ellaway A. Ecological approaches: rediscovering the role of the physical and social environment. In: Berkman LF, Kawachi I, eds. *Social epidemiology*. New York: Oxford University Press, 2000:332–48.
- 17 **Blakely TA**, Woodward AJ. Ecological effects in multi-level studies. *J Epidemiol Community Health* 2000;**54**:367–74.
- 18 **Boyle MH**, Willms JD. Place effects for areas defined by administrative boundaries. *Am J Epidemiol* 1999;**149**:577–85.

THE JECH GALLERY.....

El Salvador: social cost of neoliberal health reform

Despite eight months on strike against health privatisation,¹ popular marches, and a large electoral reverse, the Salvadorian government seems determined on privatisation. The Congress debated a loan from the Interamerican Development Bank to decentralise the Ministry of Public Health (MPH), which was denounced as concealed privatisation by political opposition. Compared with social security reduced cover (15%), MPH provides services particularly to the poor and extremely poor groups. Its privatisation would enlarge the already disconcerting exclusion of these Salvadorians, condemned by the system to consume the unique item that remains free and abundant in this country—waste.

(Translation of the sign “Well done Salvadorian . . . Consume yours with pride” campaign sponsored by Salvadorian Association of Industrialists, ASI.)

E A Espinoza

Universidad de El Salvador, Final 25 Avenida Norte y Boulevard de Los Héroes Edificio de la Rectoría San Salvador, El Salvador; espinoza@telesal.net

- 1 **Espinoza E**, Vargas F. On the right to health in El Salvador: no to privatisation. *J Epidemiol Community Health* 2003;**57**:82.

