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

Published: 2003-01-01

Citation for published version (APA):
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)2,3 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).4,5 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.∗ 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.

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

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, 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).6,7 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.8 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.
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 simple because they reside in apparently middle and low risk neighbourhoods. 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.

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). 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 intra-class correlation are not contraindutive facts, but they give different and complementary information.

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 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.

CONCLUSIONS

Certainly broad social and economic forces generate differences among neighbourhoods that shape the distribution of health outcomes. Strategies of disease prevention need to combine person centred approaches with approaches aimed at changing residential environments. 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. However, when it comes to evaluating multilevel risk distribution and the public health relevance of specific administrative boundaries (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 Göllberg and Klaus Larsen for revising and commenting on the manuscript.

REFERENCES

THE JECH GALLERY
El Salvador: social cost of neoliberal health reform

Despite eight months on strike against health privatisation, popular marches, and a large electoral reverse, the Salvadoran 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