Int J Public Health 54 (2009) 129–130
1661-8556/09/030129-2
DOI 10.1007/s00038-009-7075-z

Editorial

Multilevel analysis of survey data

Herman van Oyen

Herman van Oyen is Head of the Unit of Epidemiology at the Scientific Institute of Public Health in Brussels, Belgium and visiting Professor at the Department of Public Health, University of Ghent, Belgium.

Using the 1930 census, W.S. Robinson[1] showed that although there was a positive correlation between the literacy rate in each of the 48 states in the US and the proportion of the population born outside the US, the correlation was negative when individuals were considered. The reason for this discrepancy was that immigrants tended to settle in states where the native population was more literate. Because of possible ecological fallacy-bias, epidemiologists and public health researcher have felt hesitant about deducing conclusions about individuals using ecological data or data population-level. However, individuals in a population are more than a simple sum of a set of loners. Individuals in a population are living within an ecosystem and there is a complex interaction between the health of the individual and the contextual factors of the environment as well as personal factors.

In this issue of International Journal of Public Health, two papers are published, in which the authors combined in their analysis information of individuals with information available on populations. They used multilevel models. These types of models are also known as hierarchical models because of (1) the structure of the data (e. g. students clustered within schools) and (2) the hierarchy within the model with the parameters of the within-population regressions at the bottom, controlled by hyperparameters, the parameters of the upper-level models. In the first paper[2], an Index of Relative Socio-Economic Disadvantage score (IRDS) was used as an indicator of local area disadvantage. The effect of this contextual measure on health-related quality of life (physical, mental health indicators) and lifestyle factors was estimated after accounting for individual socio-economic factors. The authors found independent associations between local area IRDS and health-related quality of life, obesity and smoking. After adjusting for individual socio-economic factors, no independent association was seen between local area IRDS and a cardiovascular risk index, chronic lung disease symptoms, leisure time physical activity and high risk alcohol use. In the second paper[3], the potential influence of school environment, defined as the school level opportunity for physical activity, on the physical activity of nine-year old children was studied. It was examined if the association between school opportunity and leisure time physical activity was different by gender and BMI status. The authors found that the association between the physical activity frequency of the children and the school-level physical activity opportunity differs by gender and BMI and that overweight boys profit from a school policy and environment favoring physical activity.

Survey data are often collected with an inherent multilevel structure (e. g. students within school, patients within hospital or data from cluster sampling, annual series of a survey). Based on such sampling design, a more traditional approach of data analysis in both studies would have considered the varying effects of the groups (local area, school) as "noise". However, a regression with predictors at the individual and groups level would not have corrected for between groups differences. The feature that distinguishes multilevel models from classical regression is in the modeling of the variation between groups. The added value of the multilevel approach in both papers is threefold because it allows (1) to study effects that vary by group, (2) to study the group characteristics contributing to this differential effect and (3) to study individual-group interaction effects.

Although attractive, not all analyses need a multilevel approach. The use of multilevel models should be reserved for specific inferential objectives e. g. as a direct way to include indicators for clusters at all levels of the design. The multilevel approach, as both papers show, is a powerful and interesting tool to combine outcome measures at individual level and predictors at the individual and higher level.

References

1. Robinson, WS. Ecological Correlations and the Behavior of Individuals. *American Sociological Review* 1950;15:351–7. doi:10.2307/2087176

2. Adams RJ, Howard N, Tucker G, Appleton S, Taylor AW, Chittleborough C et al. Effects of area deprivation on health risks and outcomes: a multilevel, cross-sectional, Australian population study. Int J Public Health 2009;54;183

3. Barnett TA, O'Loughlin JL, Gauvin L, Paradis G, Hanley J, McGrath JJ et al. School opportunities and physical activity frequency in nine year old children. Int J Public Health 2009;54;150

Address for correspondence

Herman van Oyen
Unit of Epidemiology
Scientific Institute of Public Health
J. Wytmanstraat 14
B-1050 Brussels
Belgium
E-mail: herman.vanoyen@iph.fgov.be