THEORY AND METHODS

A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people

Juan Merlo, Min Yang, Basile Chaix, John Lynch, Lennart Råstam

Study objective: (1) To provide a didactic and conceptual (rather than mathematical) link between multilevel regression analysis (MLRA) and social epidemiological concepts. (2) To develop an epidemiological vision of MLRA focused on measures of health variation and clustering of individual health status within areas, which is useful to operationalise the notion of “contextual phenomenon”. The paper shows how to investigate (1) whether there is clustering within neighbourhoods, (2) to which extent neighbourhood level differences are explained by the individual composition of the neighbourhoods, (3) whether the contextual phenomenon differs in magnitude for different groups of people, and whether neighbourhood context modifies individual level associations, and (4) whether variations in health status are dependent on individual level characteristics.

Design and participants: Simulated data are used on systolic blood pressure (SBP), age, body mass index (BMI), and antihypertensive medication (AHM) ascribed to 25,000 subjects in 39 neighbourhoods of an imaginary city. The paper concentrated on SBP variance between individuals and neighbourhoods as a function of individual BMI.

Results: The variance partition coefficient (VPC) showed that clustering of SBP within neighbourhoods was greater for people with a higher BMI. The composition of the neighbourhoods with respect to age, AHM use, and BMI explained about one fourth of the neighbour differences in SBP. Neighbourhood context modified the individual level association between BMI and SBP. Individual level differences in SBP within neighbourhoods were larger for people with a higher BMI.

Conclusions: Statistical measures of multilevel variations can effectively quantify contextual effects in different groups of people, which is a relevant issue for understanding health inequalities.

Persons with similar characteristics may have different health experiences according to whether they live in one neighbourhood or in another, and people living in the same neighbourhood tend to experience a similar health status. This is in part because people in the same neighbourhood are subject to common contextual influences. This contextual phenomenon expresses itself as clustering of individual health status within neighbourhoods and is a core notion in social epidemiology. It is necessary to use procedures like multilevel regression analysis (MLRA) to consider the connection between the outcome health variable and individual level characteristics. The clustering of individual health within areas, however, may reflect the varying composition of neighbourhoods in terms of individual characteristics, rather than true contextual effects. Therefore, we also show how to determine the extent to which neighbourhood disparities in health may change after taking into account differences in the individual composition of the neighbourhoods. Finally, we illustrate that the neighbourhood context may modify the association between individual characteristics and health. The epidemiological approach to MLRA used here is of relevance in public health for understanding both the patterns and determinants of geographical and social disparities in health.

Our objective was to provide a didactic and conceptual (rather than mathematical) link between MLRA and social epidemiological concepts. Readers who wish a more formal or alternative statistical explanation may consult the referenced studies on multilevel analysis published elsewhere.

This report is based on hypothetical data regarding systolic blood pressure (SBP). Its conclusions, therefore, must not be taken as empirical evidence.

POPULATION AND METHODS
The study model population and variables

The population of our study model consists of 25,000 people 35 to 64 years old, living in the 39 neighbourhoods of a hypothetical city. It presents a clear multilevel structure with

Abbreviations: MLRA, multilevel regression analysis; SBP, systolic blood pressure; AHM, antihypertensive medication; VPC, variance partition coefficient; BMI, body mass index
people (level 1) nested within neighbourhoods (level 2). The individual outcome variable is SBP. For didactic reasons and to keep the model as simple as possible, only three individual variables were included: (1) age in years, (2) body mass index (BMI) in kg/m² (both of the aforementioned variables were centred on their overall mean of 49 years and 25 kg/m²), and (3) antihypertensive medication.

Table 1
Hypothetical multilevel regression analysis of systolic blood pressure (SBP) (mm Hg) in 25,000 people aged 35 to 64 from 39 neighbourhoods of a city

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Empty model</th>
<th>Model with individual variables</th>
<th>Model with individual variables and random slopes</th>
<th>Model with individual variables, random slopes, and non-constant individual variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SBP of the city (intercept)</td>
<td>130.2</td>
<td>128.1</td>
<td>128.0</td>
<td>127.9</td>
</tr>
<tr>
<td>Regression coefficients (b), 95% confidence intervals (CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age in years</td>
<td>0.81 (0.78, 0.87)</td>
<td>0.80 (0.77, 0.86)</td>
<td>0.89 (0.83, 1.01)</td>
<td>0.91 (0.85, 1.03)</td>
</tr>
<tr>
<td>BMI (1 unit kg/m²)</td>
<td>0.88 (0.86, 0.92)</td>
<td>0.89 (0.83, 1.01)</td>
<td>0.91 (0.85, 1.03)</td>
<td>0.91 (0.85, 1.03)</td>
</tr>
</tbody>
</table>

Random effects

<table>
<thead>
<tr>
<th>Components of variance</th>
<th>Variance (SE)</th>
<th>Variance (SE)</th>
<th>Variance (SE)*</th>
<th>Variance (SE)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance between neighbourhood intercepts</td>
<td>36.2 (8.1)</td>
<td>27.6 (6.4)</td>
<td>27.6 (6.4)</td>
<td>27.5 (6.4)</td>
</tr>
<tr>
<td>BMI-SBP slope variance between neighbourhoods</td>
<td>433.4 (3.9)</td>
<td>307.6 (2.8)</td>
<td>305.1 (2.8)</td>
<td>298.5 (3.5)</td>
</tr>
<tr>
<td>BMI related individual SBP variance</td>
<td>0.93 (0.34)</td>
<td>0.92 (0.34)</td>
<td>0.92 (0.34)</td>
<td>0.16 (0.08)</td>
</tr>
<tr>
<td>Individual level covariance</td>
<td>5.16 (0.31)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Proportional change in variance (PCV) by the new model

<table>
<thead>
<tr>
<th>Between neighbourhoods</th>
<th>PCV</th>
<th>PCV</th>
<th>PCV*</th>
<th>PCV*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>24%</td>
<td>24%</td>
<td>24%</td>
<td>24%</td>
</tr>
<tr>
<td>ICC-VPC</td>
<td>ICC = 0.08</td>
<td>ICC = 0.08</td>
<td>VPC = 0.08*</td>
<td>VPC = 0.08*</td>
</tr>
<tr>
<td>Deviance</td>
<td>222 764 (Reference)</td>
<td>211 585</td>
<td>211 443</td>
<td>211 058</td>
</tr>
</tbody>
</table>

Note: Fixed effects” and “random effects” are expressions that are often used in MLRA. In very simple terms, fixed effects are used to model means, whereas random effects are used to model variances. BMI, body mass index; ICC, intraclass correlation; VPC, variance partition coefficient. *As the variance is a function of individual BMI, the values in the table are for the intercepts (people with BMI equal to 25 kg/m²).

When performing statistical modelling, you try to find a function that predicts SBP as well as possible and, therefore, decrease unexplained variance. The goodness of this fit is measured by different statistical techniques. One very common technique that we used in our study model is the reduction in “deviance”. This technique is used to evaluate the fit of consecutive models with additional terms.

Table 1 shows that in our study model, as compared with the empty model, every consecutive model significantly decreases the deviance and improves the goodness of the fit of the model.

Figure 1: The regression lines represent the association between body mass index (BMI) and systolic blood pressure (SBP) in five hypothetical neighbourhoods of an imaginary city. Note that the lines are parallel. The neighbourhood lines have a different intercept but a similar slope. Therefore, differences between the neighbourhoods are constant for every BMI value (that is, the differences between points A, B, C, D, and E and between F, G, H, I, and J on the vertical dotted lines at BMI values 25 and 35 kg/m² are of similar magnitude). Note that the BMI value of 25 kg/m² corresponds to the mean BMI for the city and also corresponds to the value 0 for the centred BMI, as BMI is centred on its mean in the regression models.
respectively), and (3) whether any antihypertensive medica-
tion (AHM) was used. No neighbour variable was
investigated in this article. For simplicity, age and BMI
effects were assumed to be linear.

Our study model was adapted from an actual empirical
investigation analysing countries rather than neighbour-
hoods that was published elsewhere.14

MLRA
Using MLRA we investigate (1) whether there is significant
clustering within neighbourhoods (model 1), (2) to which
extent neighbourhood level differences are explained by the
individual composition of the neighbourhoods (model 2), (3)
whether the contextual phenomenon differs in magnitude for
different groups of people, and whether neighbour context modifies individual level associations (model 3), and (4) whether variations in health status are dependent on
individual level characteristics (model 4)

The ‘‘empty’’ MLRA model (model 1)
The empty MLRA model is more extensively discussed in a
companion paper15 and elsewhere.6–8 It does not include
explanatory variables but only estimates the city SBP mean and
the neighbourhood level and individual differences in SBP. In this model, the variations in SBP between
neighbourhoods are assumed to be of similar magnitude for
every person. In using this model, we simply aim to identify a
possible contextual phenomenon that can be quantified by
clustering of SBP within neighbourhoods.15

The MLRA model with individual variables (model 2)
This model expands the empty model (model 1) by including
the individual variables age, BMI, and use of AHM as fixed
effects. “Fixed effects” and “random effects” are expressions
that are often used in MLRA. Essentially, fixed effects are
used to model averages (for example, means or regression
coefficients), whereas random effects are used to model
differences (for example, neighbourhood variance).17

In equation 1 we present model 2 using non-conventional
notation that will be better understood by readers not trained
to read formal statistical notation. An individual level
variable is shown by subscript I, and neighbourhood
information shown by N. Readers who wish a more formal
notation that will be better understood by readers not trained
to read formal statistical notation are referred elsewhere.6–8

\[
SBP_i = SBPC + \beta_1 \times age_i + \beta_2 \times BMI_i + \beta_3 \times AHM_i + (EN_i + E_i) \\
(1)
\]

\[ SBP_i = \text{SBP of a given individual (I) in a given neighbour-
hood.} \]

Fixed effects

\[ SBPC = \text{SBP in people not on AHM and with average age}
and BMI (which correspond with a value of 0 as these
variables are centred), and living in a neighbourhood with
a shrunken residual equal to 0}^{15}\]

\[ \beta_1 = \text{regression coefficient of the association between age}
and SBP}^{15}\]

\[ \beta_2 = \text{regression coefficient of the association between BMI}
and SBP}^{15}\]

\[ \beta_3 = \text{regression coefficient of the association between AHM}
and SBP}^{15}\]

Random effects

\[ EN_i = \text{neighbourhood shrunken residual}^{15}\]

\[ E_i = \text{individual residual} \]

In MLRA models, it is assumed that the residuals are
normally distributed and that the individual level residuals
and the neighbourhood level residuals are independent. In
MLRA, both people and neighbourhoods are assumed to be
randomly sampled from a population of individuals and a
population of neighbourhoods.

As in simple regression analysis, the regression coefficients
of the individual variables in MLRA describe the association
between the individual level variables and SBP. The
regression coefficients in table 1 show that SBP increases
with age and BMI, as expected. Use of AHM is associated
with higher SBP, reflecting that the condition underlying this
treatment is simply hypertension.

The presence of neighbourhood and individual residuals in
the model reflects that SBP varies at both levels. The variance
is a summary of the differences, and including individual
variables in the statistical models allows us to obtain values
of variances that are adjusted for age, BMI, and AHM use.
The adjusted total SBP variance between people in the city
\((V_{\text{total}})\) can be partitioned into a variance between
neighbourhoods \((V_N)\) and a variance between people within
neighbourhoods \((V_I)\) as explained more extensively in an
earlier related paper.15

In both models 1 and 2, the neighbourhood SBP dif-
fers correspond to the neighbourhood variance of the
intercept. In model 2, the intercept is the value of the out-
come variable SBP when the explanatory variables are equal
to zero. In that case, the expected intercept value is equal to
\(SBPC + E_N\), which is the neighbourhood SBP estimated mean
for 49 year old people with a BMI of 25 kg/m\(^2\) (as continuous
individual level variables are centred on their means) without
AHM treatment. However, the neighbourhood variance in
model 2 is, in fact, the same for all the individuals, whatever
their individual characteristics. This aspect can be seen
graphically in figure 1, and can also be interpreted in saying
that the relation between BMI and SBP is considered to be
the same in all neighbourhoods. Using the coefficient
estimates from table 1, the expected value of SBP increases
by 0.88 mm Hg for every unit of increase in BMI, regardless
of whether the person resides in a neighbourhood or another.

Proportional change in variance at different levels
Neighbourhood differences in mean SBP may be attributable
to contextual influences or to differences in the individual
composition of neighbourhoods in terms of age, BMI, AHM
use, and other individual characteristics not considered in
our didactic study model. By adjusting for individual
characteristics in model 2, we take into account some part
of the compositional differences and explain some of the
neighbourhood variance detected in the empty model
(model 1).† The equation for the proportional change in
neighbourhood variance \((\text{PCV}_N)\) is:

\[ \text{PCV}_N = \frac{V_N}{V_{\text{total}}} \]

\* In MLRA, however, the mathematical interpretation of the regression
coefficients is not exactly the same as in the standard non-multilevel
model not adjusted for the neighbourhood residuals. Interested readers
can obtain an extended explanation elsewhere.8

† Note however that individual characteristics may be in the causal
pathway between neighbourhood characteristics and individual differ-
ences in SBP, so including individual characteristics in the model
may result in understating the contribution of contextual influences to SBP.
The interpretation of the PCV therefore depends on the individual variables
included in the model, and on their hypothesised role (that is, coresponding
factor, mediating role).

‡ The proportional change in variance is often referred to as “explained
variance”. However, the addition of individual variables in the model
may increase the second level variance. Indeed, in cases in which the
neighbourhood differences are hidden by their individual composition,
the total variance may decrease but the neighbourhood component of
the variance increase. Therefore, “proportional change in the variance”
is a more appropriate term than “explained variance”.15
where VN-1 is the neighbourhood variance in the empty model and VN-2 is the neighbourhood variance in the model including individual characteristics. For example, comparing model 1 with model 2, 

\[
PCVN = \frac{(36.2 - 27.6)}{36.2}
\]

We conclude that 24% of the neighbourhood SBP variance in the empty model was attributable to the three compositional factors considered.

This equation can be adapted to calculate the proportional change of variance at the individual (I) level (PCV_I), as individual SBP variance within neighbourhoods will also be explained by differences in age, BMI, and AHM use.

In table 1, we see that 29% of the individual SBP differences (that is, within neighbourhoods variance) in the empty model was attributable to differences in age, BMI, and use of AHM.

The intraclass correlation

Using the values of the adjusted variance at both levels we calculated the adjusted intraclass correlation (ICC_{adj}). We have explained the concept of ICC in a previous paper on this topic. This measure is of relevance, as it quantifies clustering of individual SBP within neighbourhoods and, therefore, can be used to operationalise the concept of contextual phenomena.

\[
PCV_I = \frac{(V_{I-1} - V_{I-2})}{V_{I-2}}
\]
by the same amount proportionally, what was reflected in the IPC. Predictors reduced the individual and neighbourhood level variances in SBP. The inclusion of individual level variables and random slopes (model 3) resulted in a decrease in the contribution of contextual influences to SBP. The interpretation of the multilevel regression model with individual variables and random slopes (model 3) suggests that the neighbourhood context conditions the association between BMI and SBP. This is supported by a graphic illustration of random slopes presented in figure 2, where it can be seen that the slope of BMI and SBP is steeper in some neighbourhoods than in others.

The intercept variance would be the neighbourhood variance for people with a BMI equal to 0 if BMI was not centred. The intercept variance: Neighbourhood variance for people with centred BMI equal to 0 kg/m² that corresponds with a non-centred BMI of 25 kg/m². The figure represents a case in which the intercept variance is zero (on the dotted line at BMI equal to 25 kg/m²). However, because of random slopes, even if the variance is zero for the intercept, the variance is very large for people with a BMI of 35 kg/m² (the dotted line at BMI equal to 35 kg/m²). Centring the variables on the city BMI mean ensures that the neighbourhood intercept variance can be directly interpreted. This figure shows that, if the BMI was not centred on its mean of 25 kg/m², the intercept variance would correspond to people with a BMI equal to 0 kg/m², who obviously do not exist. The regression lines have been prolonged as dotted lines to represent the intercept variance if the BMI was not centred.

\[
\text{ICC}_{\text{Adj}} = \frac{V_{N-2}}{V_{N-2} + V_{I-2}} \quad (4)
\]

The ICC_{Adj} is the proportion of total variance in SBP that remains at the neighbourhood level after taking into account the individual composition of the neighbourhoods in terms of age, BMI, and AHM use. Table 1 shows that about 8% (ICC equal to 0.08) of the individual residual differences in SBP were related to the neighbourhood level and might be attributable to contextual factors. Even if compositional confounding remains in the data, model 2 suggests that the neighbourhood context conditions a common level of blood pressure over and above individual age, BMI, and AHM use.

Note that in models 1 and 2, we have calculated a single ICC value for the whole city, with the assumption that all people are influenced by the neighbourhood context in an equivalent extent.

The multilevel regression model with individual variables and random slopes (model 3)

In model 3 (equation 5) we relinquish the idea that the magnitude of the association between BMI and SBP is similar in all neighbourhoods. Rather, we assume that the effect of BMI on SBP may vary depending on the neighbourhood context. For example, it is possible that contextual factors in some neighbourhoods have a higher impact on overweight people than in people with normal BMI. In model 3, therefore, we extend model 2 by allowing the regression coefficient of BMI to vary randomly at the neighbourhood level:

\[
\beta_{2C} = \text{mean regression coefficient of the association between BMI and SBP in the city}
\]

\[
\beta_{N-C} = \text{shrunken difference between } \beta_{2C} \text{ and the specific regression coefficient in a given neighbourhood}
\]

\[
E_{N-BMI} = \text{shrunken difference between the city SBP mean and the neighbourhood SBP mean for individuals with a BMI equal to 25. See our previous related paper for an explanation of the concept of shrunken residual.}
\]

Neighbourhood mean and slope differences (intercept and slope variances between neighbourhoods)

In model 3, aside from the differences in mean SBP (that is, the intercept variance V_{N-C}), each neighbourhood has its own regression coefficient for the association between BMI and SBP, and each neighbourhood coefficient deviates from the city mean regression coefficient (\beta_{2C}) by a residual amount (E_{N-BMI}). This slope variability is measured by the slope variance (V_{N-BMI}). In MLRA, this procedure is called random slope analysis or random cross level interaction analysis. It suggests that the neighbourhood context modifies the individual level association between BMI and SBP. A graphic illustration of random slopes is presented in figure 2, where it may be seen that the slope of the association between BMI and SBP is steeper in some neighbourhoods than in others.

Correlation between the intercept and slope residuals

It may occur that in those neighbourhoods with a high mean SBP (a high intercept residual value), the slope of the association between BMI and SBP is steeper (the residual value for the slope is higher). We can express this connection as a coefficient of correlation according to equation 6:

\[
\text{Correlation} = \frac{\text{COV}_{N-C}(N-BMI)}{\sqrt{V_{N-C}} \times V_{N-BMI}} \quad (6)
\]
where the variances \((V_{N-c}, V_{N-BMI})\) and the covariance between the intercept and slope residuals \((\text{Cov}(N-c, N-BMI))\) are parameters that are directly estimated by the multilevel model.

Table 1 shows that the slope variance \((V_{N-BMI})\) is 0.11, the intercept variance \((V_{N-c})\) is 27.6, and the covariance between the intercept and the slope residuals \((\text{Cov}(N-c, N-BMI))\) is 0.93. Using the estimates and applying equation 6 we find that the correlation between the intercept and the slope is equal to 0.53. Figure 3 illustrates this correlation and suggests that on average BMI has a stronger impact on SBP (that is, the slope of the association is steeper) in neighbourhoods with a higher mean level of SBP (that is, a higher value of the intercept).

Beyond the interest of examining whether an individual level association varies between neighbourhoods, a relevant reason to consider random slopes is to examine whether neighbourhood differences in SBP have a different magnitude among people with different characteristics. Figures 2 and 4 give a graphic illustration of this concept. In our study model the neighbourhood variance now depends on individual BMI; it is no longer a single value as in models 1 and 2, but a function of BMI. The figure 2 shows that there is more neighbourhood variability in SBP among those who have a higher BMI.

Considering neighbourhood variance as a function of individual level variables (equation 7) yields improved information on the shape of neighbourhood differences. The neighbourhood variance function and its confidence intervals are directly calculable by available software. The interested readers will find another practical example and a more formal statistical explanation elsewhere.

\[
V_N = V_{N-c} + 2 \times \text{Cov}(N-c, N-BMI) \times \text{BMI} + V_{N-BMI} \times \text{BMI}^2
\]  

\[\text{model 3} (\text{equation 5})\]

Table 1 provides the values of the intercept variance, the slope variance, and the covariance between the intercept and the slope needed in equation 7. The full shape of the neighbourhood differences as a function of BMI is presented in figure 5A. These neighbourhood differences seem to be larger for overweight people, reflecting the fact that the neighbourhood context modifies the individual association between BMI and SBP.

Note that in models 1 and 2 with random intercept only, the neighbourhood level variance is assumed to be the same for all people. We can, therefore, compare the intercept variance in models 1 and 2 with any value of the variance function \(V_N\) in model 3.

It is important to note that the assumption that neighbourhood differences are the same for all people may conceal significant contextual effects that pertain to specific groups of people. In such cases, the neighbourhood heterogeneity can only be properly identified when random slopes are taken into account and the variance is calculated as a function of individual characteristics. Figure 4 contains a hypothetical situation in which the neighbourhood context strongly modifies the individual association between BMI and SBP. In figure 4, the neighbourhood differences (intercept variance) is close to null; but for people whose centred BMI is over and under zero, the neighbourhood variance is very large and suggests that the context has apparent importance.

The multilevel regression model with individual variables, random slopes, and non-constant individual variance (model 4)

We can hypothesise that within each neighbourhood individual SBP differences are higher among those who are overweight than among people of normal BMI. In statistical terms this phenomenon is called individual level heteroscedasticity, meaning that the individual level variance in SBP is not constant along BMI. This is shown in figure 2, where triangles surrounding the regression lines show that individual level SBP variations within neighbourhoods increase with BMI. Absence of heteroscedasticity is a precondition for doing correct regression analysis. We can use MLRA to model non-constant individual level variance and obtain both relevant epidemiological information and correct regression estimates.

Model 4 (equation 8) is similar to model 3 (equation 5) but includes an additional individual level residual, which is related to BMI \((E_{I,BMI})\).
\[
SBP_i = (SBP_{ci} + E_{i-\epsilon}) + \beta_1 \times AGE_i \\
+ (\beta_2C + E_{i-\epsilon}(1-BMI)) \times BMI_i + \beta_3 \times AHI_i \\
+ (E_{i-\epsilon} + E_{i-\epsilon}(1-BMI)) \times BMI_i
\]

As every group of people with a particular BMI value has a specific individual level SBP variance, the individual level variance also becomes a function of BMI.

In equation 9 we provide a very simple expression of this function; more details on the equation may be found elsewhere. As in the case of neighbourhood variance, this variance function is directly calculable with available softwares.

\[
V_1 = V_{1-\epsilon} + 2 \text{COV}(1-\epsilon(1-BMI)) \times BMI_i + V_{1-BMI}^2
\]

where \( V_{1-\epsilon} \) is the individual variance related to the intercept, \( V_{1-BMI} \) is the individual level variance related to BMI, and \( \text{COV}(1-\epsilon(1-BMI)) \) is the covariance between the two sets of individual level residuals.

Modelling of individual variance may not only significantly improve the fit of the statistical model and the validity of the regression estimates but also provides useful information when it comes to understanding SBP inequalities among individuals and neighbourhoods.

Variance functions and the variance partition coefficient (VPC)

In contrast with models 1 and 2, the neighbourhood level and individual level SBP variances in models 3 and 4 are no longer represented by one simple value. The complete picture of the variances in our study model is presented in figure 5A that has been obtained using equations 7 and 9. Figure 5A shows that both individual level and neighbourhood level variances increase considerably with BMI.

As suggested earlier, contextual influences may be stronger for certain groups of people such as overweight people, and less important for people with normal BMI. To quantify this aspect, we must examine how differences in SBP are partitioned between the individual level and the neighbourhood level for different BMI values. Rather than one ICC, we calculate a VPC that is function of the BMI (VPCBMI), using the neighbourhood level (\( V_N \)) and individual level (\( V_I \)) variance functions of equations 7 and 9.

\[
\text{VPC}_{\text{BMI}} = \frac{V_N}{V_N + V_I}
\]

As shown in table 1, the VPC in model 3 is about 0.08 in people with a BMI of 25 kg/m², which means that 8% of the variations for these people were at the neighbourhood level. The VPC as a function of BMI is presented in figure 5B and shows that the VPC was equal to 0.14 (14%) for people with a BMI of 40 kg/m². Even if the values of variance function at the extremes of the curve are less reliable, the neighbourhood context seems to play a more relevant part for people with high BMI.

**DISCUSSION**

We have attempted, on the basis of hypothetical data, to illustrate the investigation of a contextual phenomenon that expressed itself as a clustering of individual SBP within neighbourhoods. According to our findings, this contextual phenomenon was—at least partly—indeed of the composition of the neighbourhoods, and has a different impact on people with a different BMI. Indeed, as people living in the same neighbourhood share common contextual influences, they tend to experience a similar SBP level. For example, you might hypothesise that neighbourhood disparities in access and quality of hypertension care may condition neighbourhood specific degrees of blood pressure control over and above individual differences.

In our study model we found that the neighbourhood context influenced overweight people to a greater extent than normal weight people. Possible explanations could be that physicians in some neighbourhoods treat overweight people with hypertension more intensively than physicians in other neighbourhoods, or that interventions directed to factors that affect SBP (for example, physical activity or low salt diet) are more efficient in certain neighbourhoods.

The association between BMI and SBP was stronger in neighbourhoods with a higher level of SBP (the correlation between the intercept and the slope was equal to 0.53). To explain this pattern, it could be hypothesised that obese people tend to present higher SBP in neighbourhoods with less successful blood pressure control strategies (that is, with higher SBP means).

Therefore, as a whole, our study emphasises a vision of MLRA that focuses on measures of health variation for understanding the distribution of health status in the population. In describing complex patterns of variation, MLRA provides useful information for analysis of cross level (for example, neighbourhood-individual) causal pathways.

Despite their relevance, some of the concepts presented here have not been widely discussed in the literature on multilevel analysis. The analysis of patterns of variance, which has been undervalued in many previous investigations, contributes to our understanding of the distribution and determinants of geographical, social, and individual disparities in health status.

Our didactic presentation shows the link between the statistical concepts of MLRA and the social epidemiological notion of contextual phenomenon and permits a better assessment of the interest of MLRA in social medicine and public health research.

**Authors’ affiliations**

J Merlo, L Råstam, Department of Community Medicine (Preventive Medicine), Malmö University Hospital, Lund University, Malmö, Sweden
M Yang, Institute of Community Health Sciences, Queen Mary University of London, London, UK
B Chaux, Research Team on the Social Determinants of Health and Healthcare, National Institute of Health and Medical Research, Paris, France
J Lynch, Department of Epidemiology, Center for Social Epidemiology and Population Health, University of Michigan, Ann Arbor, Michigan, USA

Funding: this study is supported by grants (principal investigator Juan Merlo) from the Swedish Council for Working Life and Social Research (number 2002-054 and number 2003-0580), and from the Swedish Research Council (number 2004-6155).

Conflicts of interest: none.

**REFERENCES**

In addressing the moral consequences of classification, Bowker and Star have remarked that “each category values some point of view and silences another”.[1] Unusually in medicine, such valorisation has silently entered the Read Codes with the announcement by the NHS Centre for Coding and Classification in January 2004 of the addition of “136R. Binge drinker”. Medical terms usually qualify a condition or disease rather than label the person who has it. The reason is that personal labels may carry negative connotations, arousing resistance or defensiveness, whereas terms for behaviours (such as “high risk drinking”, “alcohol dependence”, and even “binge drinking”) are more likely to vary, too. It was defined in a report of the Royal College of Physicians as “a man who regularly drinks 10 or more units in a single session, or a woman who regularly drinks 7 or more units in a single session”.[4] The new national alcohol harm reduction strategy uses the definition of eight or more units in a single session for men and six or more for women.[5]

Quotidian practice has tied “binge drinker” strongly to negative stereotypical images of the young (and frequently female, although statistically more likely to be male). The attachment of this category to a person is not a neutral process: it entails assignment rather than self ascription, a degree of visibility to others, and the likelihood of durability. Furthermore, the coupling is under the control of the medical establishment with respect to reversibility. There must be better terms in the drinking repertoire for describing this behaviour that avoid labelling people.

Correspondence to: Mr Peter J Aspinall, University of Kent, Centre for Health Services Studies, George Allen Wing, Canterbury CT2 7NF, UK; p.j.aspinall@kent.ac.uk

REFERENCES
A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people
Juan Merlo, Min Yang, Basile Chaix, John Lynch and Lennart Råstam

J Epidemiol Community Health 2005 59: 729-736
doi: 10.1136/jech.2004.023929

Updated information and services can be found at:
http://jech.bmj.com/content/59/9/729

These include:
References
This article cites 9 articles, 7 of which you can access for free at:
http://jech.bmj.com/content/59/9/729#BIBL
Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
Health service research (832)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/