

## THEORY AND METHODS

# A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health

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

*J Epidemiol Community Health* 2005;59:1022–1029. doi: 10.1136/jech.2004.028035

See end of article for authors' affiliations

Correspondence to: Professor J Merlo, Department of Community Medicine (Section of Preventive Medicine), Malmö University Hospital, Faculty of Medicine (Campus Malmö), Lund University, S-205 02 Malmö, Sweden; juan.merlo@med.lu.se

Accepted for publication 1 December 2004

**Study objective:** Using a conceptual rather than a mathematical approach, this article proposed a link between multilevel regression analysis (MLRA) and social epidemiological concepts. It has been previously explained that the concept of clustering of individual health status within neighbourhoods is useful for operationalising contextual phenomena in social epidemiology. It has been shown that MLRA permits investigating neighbourhood disparities in health without considering any particular neighbourhood characteristic but only information on the neighbourhood to which each person belongs. This article illustrates how to analyse cross level (neighbourhood–individual) interactions, how to investigate associations between neighbourhood characteristics and individual health, and how to use the concept of clustering when interpreting those associations and geographical differences in health.

**Design and participants:** A MLRA was performed using hypothetical data pertaining to systolic blood pressure (SBP) from 25 000 subjects living in the 39 neighbourhoods of an imaginary city. Associations between individual characteristics (age, body mass index (BMI), use of antihypertensive drug, income) or neighbourhood characteristic (neighbourhood income) and SBP were analysed.

**Results:** About 8% of the individual differences in SBP were located at the neighbourhood level. SBP disparities and clustering of individual SBP within neighbourhoods increased along individual BMI. Neighbourhood low income was associated with increased SBP over and above the effect of individual characteristics, and explained 22% of the neighbourhood differences in SBP among people of normal BMI. This neighbourhood income effect was more intense in overweight people.

**Conclusions:** Measures of variance are relevant to understanding geographical and individual disparities in health, and complement the information conveyed by measures of association between neighbourhood characteristics and health.

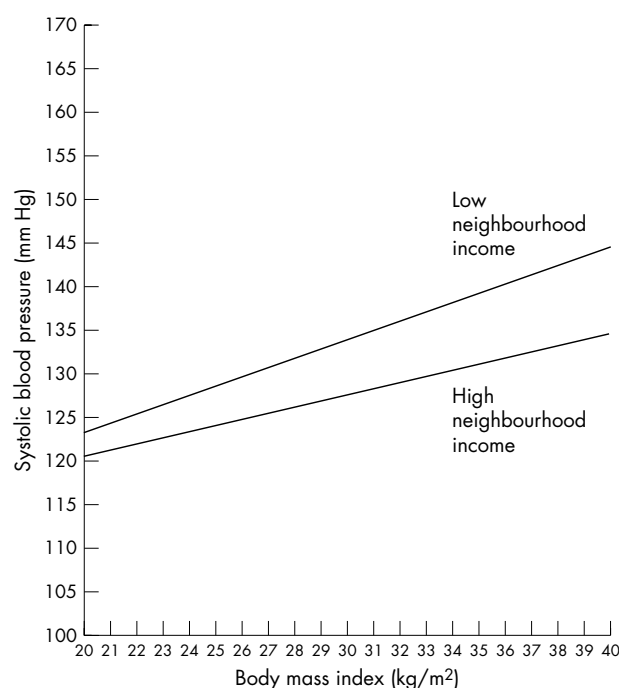
An aspect of central importance in social epidemiology is that the health differences between people can partly be attributed to the areas in which they live.<sup>1–3</sup> People with similar characteristics who live in different neighbourhoods may have different health statuses because of differing cultural, economic, political, historical, or geographical influences. In other words, different people may to some extent share similar health statuses because they share a common environment.<sup>4</sup> This contextual phenomenon expresses itself as the clustering of individual health status within neighbourhoods.<sup>5–7</sup> The presence of clustering is, in turn, the main reason for applying multilevel regression techniques. If the individual health status is correlated within neighbourhoods, analysis using common regression methods underestimates the standard errors for contextual effects and gives biased results. Multilevel regression analysis (MLRA) is a statistical methodology that provides information on how health disparities are distributed between the individual and the neighbourhood levels, quantifies the clustering of individual health status within neighbourhoods, and permits examining cross level interactions between the effects of neighbourhood and individual level factors.<sup>8–11</sup>

Using an hypothetical example on systolic blood pressure (SBP), we have previously explained that MLRA can be used to investigate neighbourhood disparities in health without considering any specific neighbourhood characteristic but only the neighbourhood affiliation of each person.<sup>12–13</sup> A

feature of those previous analyses was that the characteristics of the neighbourhood context remained unspecified. In MLRA, neighbourhood effects can actually be assessed by quantifying and modelling health differences, or variance, alone.

In this essay we go one step further in using MLRA to build an understanding of how this contextual phenomenon operates. We explain that neighbourhood differences in SBP can be accounted for by specific contextual characteristics such as the income level of a neighbourhood. This socioeconomic contextual characteristic may affect individual health over and above individual socioeconomic characteristics. We show that classic measures of association (for example, regression coefficients) between specific neighbourhood characteristics and health might help to elucidate possible cross level causal pathways. We also show how to investigate whether neighbourhood factors modify the effect of specific individual variables. We finally explain why those measures of association need to be interpreted together with measures of variance to improve our understanding of geographical health disparities.

**Abbreviations:** MLRA, multilevel regression analysis; SBP, systolic blood pressure; BMI, body mass index; AHD, antihypertensive drug; MLRA, multilevel regression analysis; VPC, variance partition coefficient; PCV, proportional change in variance



**Figure 1** Fixed effect cross level interaction between neighbourhood low income and individual BMI. It shows that the increase of systolic blood pressure (SBP) associated with body mass index (BMI) is steeper in deprived neighbourhoods.

## POPULATION AND METHODS

### The hypothetical example

Our hypothetical population consists of 25 000 subjects 35 to 64 years old, living in the 39 neighbourhoods of an imaginary city. There is a clear multilevel structure with individuals (level 1) nested within neighbourhoods (level 2). The individual outcome variable is SBP. For the didactic purpose only four individual variables were included: (1) age in years, (2) body mass index (BMI) in  $\text{kg/m}^2$  (these two variables have been centred on their overall means of 49 years and 25  $\text{kg/m}^2$ , respectively), (3) whether any antihypertensive drug (AHD) was used, and (4) individual low income defined as having an income under the median in the city. The neighbourhood low income was defined as having an average income under the median neighbourhood income in the city.

Our hypothetical study model was adapted from an empirical investigation, published elsewhere, analysing countries rather than neighbourhoods.<sup>14</sup>

### MLRA

In this paper we follow a sequence of analysis that we also used in two previously published articles.<sup>12–13</sup> In these articles we described an “empty” MLRA model (model i),<sup>12</sup> a model including individual variables only (model ii), a model with a random slope† for BMI (model iii), and a model further taking into account a non-constant individual variance‡ related to BMI (model iv).<sup>13</sup>

This paper completes the sequence of analysis described in previous publications<sup>12–13</sup> by including information on individual and neighbourhood socioeconomic position (that is, income), which permits disentangling their specific effects on SBP level.

\*The “empty” MLRA model (model i) is a model without any independent variable. The outcome (for example, SBP) is only function of the area in which the people live, which is accounted for with an area level random intercept. See elsewhere<sup>12</sup> for a more detailed explanation.

In model iv<sup>0</sup>, which is a simple extension of model iv described in detail in a previous companion article,<sup>13</sup> we additionally include individual low income. In model v we further introduce neighbourhood low income. In model v<sup>0</sup> we expand model v by including a term for interaction between individual BMI and neighbourhood low income effects. This last model allows us to examine whether the effect of BMI on SBP is more intense in low income neighbourhoods or, to say things in another way, whether SBP differences between low and high income neighbourhoods are larger in people with higher BMI.

### The model iv<sup>0</sup> can be described as:

$$SBP_i = (SBP_C + E_{N-C}) + \beta_1 \times Age_i + (\beta_{2C} + E_{N-BMI}) \times BMI_i + \beta_3 \times AHD_i + \beta_4 \times Low\ income + (E_{I-C} + E_{I-BMI} \times BMI_i)$$

$SBP_i$  = SBP of a given individual,  $i$ , in a given neighbourhood.

### Fixed effects§

$SBP_C$  = intercept of the regression. In this model  $SBP_C$  is the SBP value for people with value 0 for all the explanatory variables considered and with an average neighbourhood effect. See elsewhere<sup>13</sup> for a more detailed explanation.

$\beta_1$  = regression coefficient of the association between age and SBP

$\beta_{2C}$  = regression coefficient of the association between BMI and SBP. In this model  $\beta_{2C}$  represents the average association between BMI and SBP, which association is weaker in certain neighbourhoods and stronger in other neighbourhoods. See elsewhere<sup>13</sup> for a more detailed explanation.

$\beta_3$  = regression coefficient of the association between AHD and SBP

$\beta_4$  = regression coefficient of the association between low income and SBP

### Random effects¶

$E_{N-C}$  = shrunken difference¶ between  $SBP_C$  and the mean SBP of the neighbourhood  $N$

$E_{N-BMI}$  = shrunken difference between  $\beta_{2C}$  and the specific regression coefficient in a given neighbourhood

$E_{I-C}$  = individual level residual

$E_{I-BMI}$  = BMI related individual level residual

†In the model with the random slope (model iii) we assume that the effect of BMI on SBP may be different from one neighbourhood to another. In that case, the slope of the association between BMI and SBP would vary from one neighbourhood to another and neighbourhood disparities become a function of individual BMI. See elsewhere<sup>13</sup> for a more detailed explanation.

‡In the model with non-constant individual variance we assume that within each neighbourhood, individual SBP differences were larger among people who were 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. Absence of heteroscedasticity is a condition for performing correct regression analysis. We can use MLRA to model non-constant individual level variance and obtain both relevant epidemiological information and correct regression estimates.

§“Fixed effects” and “random effects” are expressions that are often used in MLRA. Essentially, fixed effects are used to model averages and refer to the usual regression coefficients, whereas random effects are used to model differences (for example, neighbourhood variance).

¶The shrunken differences or shrunken residuals can be estimated using the raw residuals, the estimated variances, and the number of people in each neighbourhood. The fewer the number of people in a neighbourhood, or the higher the variability within neighbourhoods as compared with the variability between neighbourhoods, the more important the shrinkage and the more the value of the neighbourhood residual will be shrunken towards 0. See elsewhere<sup>12</sup> for a more detailed explanation.

**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

	Model with individual variables and complex variance (iv <sup>0</sup> )	Model with neighbourhood variables (v)	Model with neighbourhood variables (v†)
<b>Fixed effects</b>			
(Traditional measures of association)			
Mean SBP of the city (intercept)	119 mm Hg β* (95% CI)	116 mm Hg β* (95% CI)	116 mm Hg β* (95% CI)
Individual level variables			
Age (years)	0.76 (0.70 to 0.82)	0.76 (0.70 to 0.82)	0.76 (0.70 to 0.82)
Use of blood pressure lowering drugs (yes v no)	0.89 (0.77 to 1.01)	0.88 (0.76 to 1.00)	1.02 (0.88 to 1.16)
BMI (kg/m <sup>2</sup> )	15.12 (14.43 to 15.81)	15.12 (14.43 to 15.81)	15.12 (14.43 to 15.81)
Individual low income (low v high)	2.27 (1.72 to 2.82)	2.27 (1.72 to 2.82)	2.26 (1.71 to 2.81)
Area level variables			
Neighbourhood low income (low v high)		5.23 (3.19 to 7.27)	4.47 (2.49 to 6.45)
Cross level interactions			
Neighbourhood low income × BMI†			0.35 (0.15 to 0.55)
<b>Random effects</b>			
(Multilevel measures of health variation)			
Components of variance	Variance‡(SE)	Variance‡(SE)	Variance‡(SE)
Neighbourhood intercept variance	27.09 (6.26)	21.11 (4.7)	21.02 (4.9)
Individual level variance	298.22 (3.34)	298.03 (3.34)	298.02 (3.34)
BMI-SBP slope variance	0.11 (0.03)	0.11 (0.03)	0.09 (0.03)
BMI related individual SBP variance	0.16 (0.08)	0.16 (0.08)	0.16 (0.08)
Neighbourhood intercept-slope covariance	0.99 (0.35)	0.62 (0.29)	0.58 (0.27)
Individual level covariance	5.24 (0.31)	5.24 (0.31)	5.25 (0.31)
Proportional change in variance (PCV)	PCV	PCV	PCV
PCV between neighbourhoods	Ref.	22%	22%
PCV between individuals	Ref.	0%	0%
Variance partition coefficient (VPC)	VPC = 0.08	VPC = 0.07	VPC = 0.07
2×log-likelihood value	211,024	211,006	210,996

Fixed effects and random effects are expressions often used in MLRA. In very simple terms, fixed effects are used to model means, whereas random effects are used to model variances. \*Regression coefficients (β) and 95% confidence intervals (95% CI). BMI, body mass index centred on the mean; PCV, proportional change in variance; VPC, variance partition coefficient. †The reference group is people living in advantaged neighbourhoods and with BMI = 25 kg/m<sup>2</sup>. ‡In people with BMI = 25 kg/m<sup>2</sup>.

**The model v can be described as:**

$$SBP_i = \text{model iv}^0 + \beta_5 \times \text{neighbourhood low income}_{-N}$$

**Fixed effects**

β<sub>5</sub> = regression coefficient of the association between neighbourhood low income<sub>-N</sub> and SBP

**The model v<sup>0</sup> can be described as:**

$$SBP_i = \text{model v} + \beta_6 (\text{neighbourhood low income}_{-N} \times \text{BMI}_i)$$

**Fixed effects**

β<sub>6</sub> = regression coefficient of the interaction between neighbourhood low income<sub>-N</sub> and BMI<sub>i</sub>

In multilevel analysis both individuals and neighbourhoods are assumed to be randomly sampled from a population of individuals and a population of neighbourhoods. In all the MLRA models presented here it is assumed that the residuals are normally distributed and that there is independence between the individual and neighbourhood residuals.

For all analyses, we used the MLwiN software version 1.2 developed by Goldstein's research group (<http://www.mlwin.com>).

**RESULTS**

**Study of fixed effects (traditional measures of association)**

The results presented in table 1 have a direct interpretation. The regression coefficient of the neighbourhood variable shows that over and above individual characteristics—including individual low income—SBP is higher in low income neighbourhoods. We can also see a significant cross level interaction between the effects of neighbourhood low income and individual BMI, so that the increase of SBP with BMI is more intense in low income neighbourhoods. Another

interpretation of this interaction is that the difference in SBP between low and high income neighbourhoods is more intense for people with a higher BMI. Figure 1 plots this interaction effect.

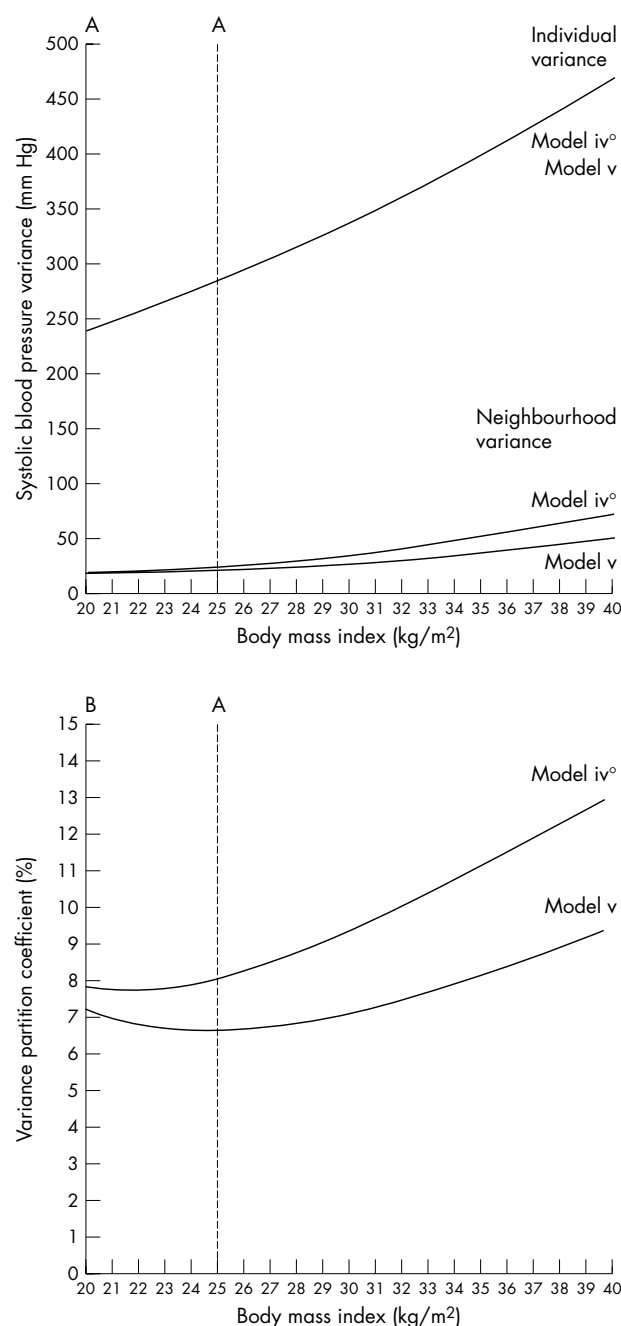
**Study of random effects (measures of health variation)**

In this section we use some general MLRA concepts, such as proportional change in variance (PCV), random slope, variance function, and variance partition coefficient (VPC) that have been presented and discussed in a previous publication.<sup>13</sup>

The neighbourhood disparities, or variance, in SBP persist after adjustment for the individual composition of the neighbourhoods in terms of age, AHD use, BMI, and individual low income. By including the neighbourhood low income variable in model v we can now observe how much of the remaining neighbourhood SBP differences are explained by the level of socioeconomic deprivation in the neighbourhoods.

We have previously explained<sup>13</sup> that the coefficient of the association between BMI and SBP—the slope—may be allowed to be random at the neighbourhood level. This procedure is called random slope analysis or random cross level interaction analysis, and informs us as to whether the effect of individual BMI on SBP differs from one neighbourhood to another. Because of the existence of a BMI slope variance, the neighbourhood variance becomes a function of BMI. The reader will find more detailed information on this aspect in a previous companion article of this series<sup>13</sup> and elsewhere.<sup>11 15</sup>

In table 1 we see that, taking the between neighbourhood variance of the intercept of model iv<sup>0</sup> as a reference of comparison, the inclusion of neighbourhood low income in model v explained 22% of the neighbourhood differences in SBP. However, as the neighbourhood variance is a function of



**Figure 2** In the upper part (A) of the figure we plot the individual level variance in systolic blood pressure (SBP) as a function of individual body mass index (BMI) and the variance in SBP at the neighbourhood level as a function of BMI. In the lower part (B) of the figure we plot the variance partitioning coefficient (VPC) defined as the percentage of the variance in individual SBP that is attributable to the neighbourhood level. The VPC is also a function of BMI. Model  $iv^0$  includes individual variables only (age, BMI, use of antihypertensive drug, and individual low income). Model  $v$  further includes the neighbourhood low income.

BMI, this percentage concerns only the intercept variance, which corresponds to the neighbourhood variance for people with a BMI of 25 kg/m<sup>2</sup> (that is, when the mean centred BMI is equal to 0). In fact, as the neighbourhood variance is a function of individual BMI, the inclusion of the neighbourhood low income in model  $v$  explains a different amount of the neighbourhood variance for people with different BMI values. This phenomenon is illustrated in figure 2A, where we see that the reduction in neighbourhood variance when

comparing model  $v$  with model  $iv^0$  is greater for people with higher BMI values. This procedure provides relevant information, showing that causal pathways between the neighbourhood socioeconomic environment and individual SBP may be differentiated for people with a different BMI.

We have previously explained that the VPC is the proportion of the variance in SBP that is located at the neighbourhood level. In the simplest form the VPC corresponds to the intraclass correlation (ICC).<sup>13</sup> In our case, the VPC is a function of BMI, as both the individual and neighbourhood variances depend on the BMI value.<sup>13</sup> It is rather intuitive that neighbourhood influences could be more intense on certain groups of individuals than on others. The VPC allows us to operationalise the idea that the clustering of a phenomenon within neighbourhoods—that is, its contextual dimension, may be of different magnitude for people with different characteristics.

We can see in figure 2B that the VPC was reduced for every person, whatever their BMI value, when information on neighbourhood low income was introduced in the model. This reduction in VPC suggests that some part of the contextual phenomenon related to individual SBP was attributable to neighbourhood deprivation. In table 1, the VPC in people with a BMI value of 25 kg/m<sup>2</sup> was slightly reduced from 0.08 in model  $iv^0$  to 0.07 in models  $v$  and  $v^0$ . However, as the between neighbourhood differences in SBP (that is, variance) depend on individual BMI values, the complete picture as to how VPC changes when the neighbourhood variable is included in the model is presented in figure 2B. Even if the interpretation of the values of the variance function at the extreme of the curve is less reliable because of the lower number of people with extreme BMI values, we realise from this figure that a larger share of the neighbourhood variations in SBP was explained for people with a higher BMI. After taking into account the neighbourhood income level, we still see that the unexplained neighbourhood disparities in SBP (that is, neighbourhood residual variance) remained more important for people with high BMI values than for people with normal BMI.

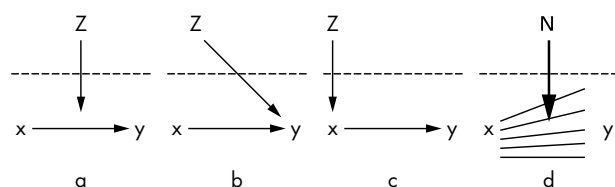
## DISCUSSION

In our hypothetical example we found that individual SBP was partly conditioned by the neighbourhood of residence of the persons. About 8% of the individual SBP differences were attributable to the neighbourhood level, after taking into account differences in the individual composition of the neighbourhoods in terms of age, use of AHD, BMI, and individual low income.

Neighbourhood low income explained a fraction of the SBP disparities, but did not account for all of the individual clustering of SBP within neighbourhoods. In other words, some part of the contextual phenomenon remained unexplained, which was expressed as a residual clustering of individual SBP within neighbourhoods.

Overall, neighbourhood low income increased individual SBP, and modified the association between BMI and SBP. Because of this interaction the effect of neighbourhood low income was especially intense among overweight people, who presented a particularly higher SBP in low than in high income neighbourhoods. A possible explanation for these hypothetical neighbourhood income effects could be that certain contextual factors, such as access to healthcare resources and other facilities, and social norms or social control that may condition blood pressure level by promoting certain lifestyles (for example, physical activity or low salt consumption), may be less favourable in low income neighbourhoods, which may be particularly detrimental for people with high BMI values in these neighbourhoods.





**Figure 3** The three important macro-micro relations also termed “cross level effects”: (a) cross level effect modification, (b) direct cross level effect and, (c) indirect cross level effect. In this figure Z is a neighbourhood level variable, x is an individual level variable, and y is an individual level outcome. These three cross level causal pathways are commonly investigated by “fixed-effect” analysis. Figures a, b, and c are adapted from a previously published work.<sup>9,16</sup> The analysis of cross level effect modifications can also be conducted with random effects (fig 3d), and examines whether the neighbourhood environment as a whole (N) modifies the association between individual variables (x–y).

### Investigating cross level effects and interactions by means of measures of association and measures of variance

We have seen the existence of a cross level (that is, neighbourhood-individual) effect on SBP. We have investigated this effect by analysing both measures of variance (random slope, variance function, PCV, and VPC)<sup>13</sup> and measures of association (regression coefficients for the association between neighbourhood low income and SBP and for the interaction between neighbourhood low income and BMI effects).

At this point, it is appropriate to briefly mention the differences between fixed and random cross level effects and cross level effect modifications. Following the conventions of Tacq discussed by Snijders<sup>9</sup> and later adapted by Blakely,<sup>16</sup> there are three important macro-micro relations also termed cross level effects: (a) cross level effect modification (that is, the interaction between neighbourhood low income and BMI effects presented in table 1), (b) direct cross level effect (that is, the direct effect of neighbourhood low income on SBP presented in table 1), and (c) indirect cross level effect. In this last case we could figure out that neighbourhood low income influences SBP by conditioning BMI which, in turn determines SBP.\*\*

The aforementioned relations are represented in figure 3 where Z is a neighbourhood level variable, x is an individual level variable, and Y is an individual level outcome. These three cross level pathways are commonly investigated by “fixed effect” analysis, that is, by studying measures of association. Relation (a) could be investigated by an interaction term ( $Z \times x$ ) or by dummy variables combining the Z and x categories. Investigating whether relations (b) or (c) exist in the data could be performed by examining whether the regression coefficient of Z changes when adjusting the model for x.

Beyond measures of association (figs 3a to 3c), cross level effect modification analysis can be conducted using random effects (fig 3d). The “random slope” analysis, rather than focusing on the effect of a specific neighbourhood variable Z on Y, examines whether the neighbourhood environment as a whole (N) modifies the association between individual variables (X–Y) without specifying any neighbourhood factors. In our hypothetical example, such cross level effect modification (that is, random slope) was described for the

\*\* Note that BMI may be in the causal pathway between neighbourhood low income and individual SBP. Whether you should adjust or not for BMI in that case depends on whether BMI is conceptualised as a confounding factor or as mediating factor. If the social environment influences health by operating as contextual determinants of other individual characteristics, then controlling for many individual factors may result in over-adjusting the true effects of the context.

relation between BMI and SBP. The reader will find more detailed information on this aspect in a previous companion article of this series<sup>13</sup> and elsewhere.<sup>11,15</sup>

### Including the concept of clustering when interpreting geographical differences in health

We have shown that measuring the clustering of individual health characteristics within neighbourhoods is a concrete way of operationalising the social epidemiological concept of contextual phenomenon.<sup>12,13</sup> Therefore, when it comes to evaluating the importance of considering neighbourhood effects on health, taking into account the size of this contextual phenomenon (that is, the size of clustering as an expression of the contextual dimension of the outcome) may be highly relevant in terms of public health.<sup>7</sup> Rather than only focusing on aggregated health differences between neighbourhoods or on associations between contextual variables and individual health, the use of measures of clustering informs us as to the relevance of between neighbourhood differences in understanding individual health disparities.

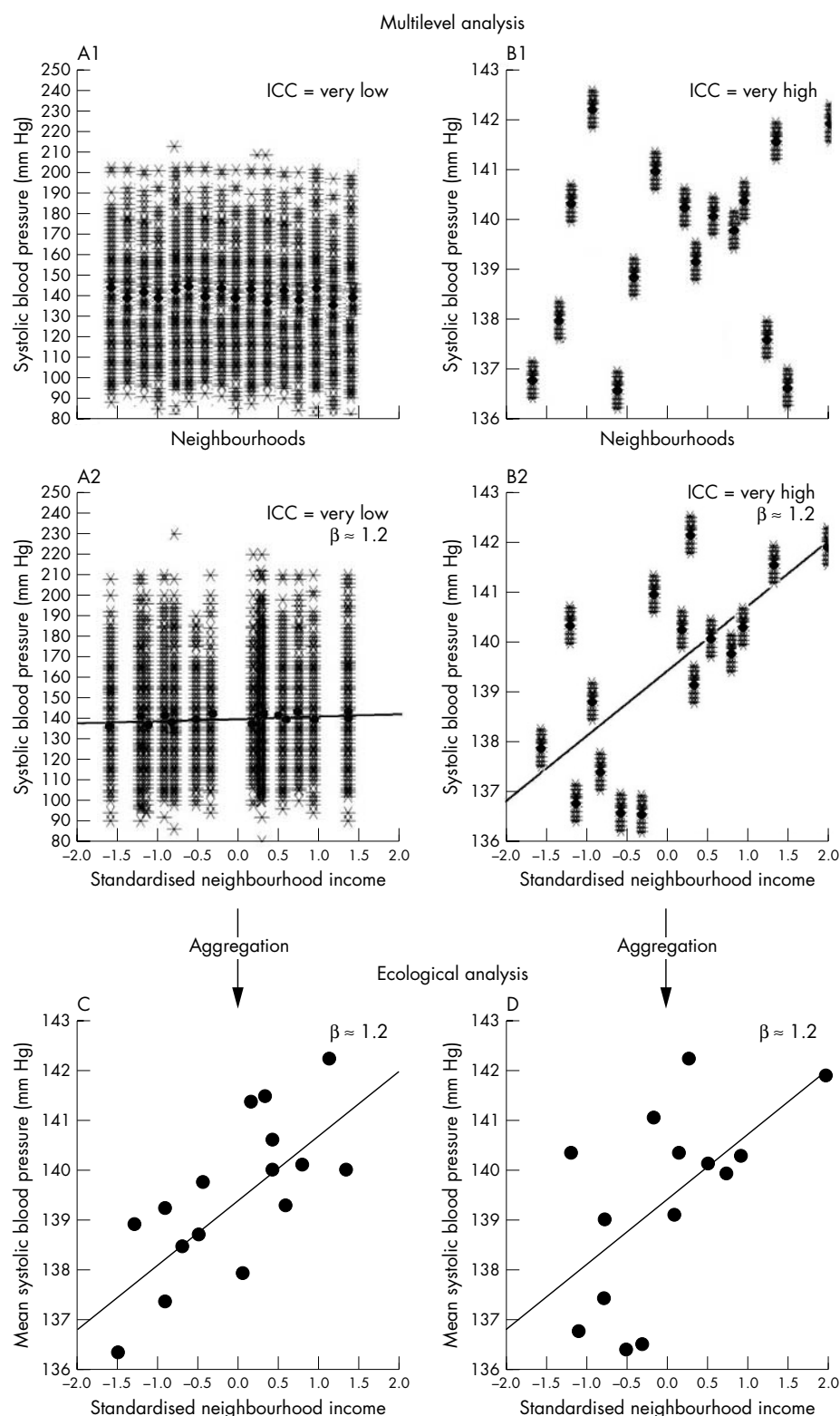
The more the health of the people within a neighbourhood is alike (as compared with the health of people from different neighbourhoods), the more probable it is that the determinants of individual health are directly related to the contextual environment of the neighbourhood, and/or that social processes of geographical segregation are taking place—that is, similar types of individuals choose or are forced to reside in a given neighbourhood.<sup>12</sup>

Figure 4 presents two sets of data showing differences in SBP (on the y axis). The upper sections of the figure (A1 and B1) illustrate the multilevel structure of the data with individual values (represented by a star) nested within neighbourhoods. In the sections A2 and B2, the neighbourhood are ordered by standardised neighbourhood mean income (the x axis). In the lower sections (C and D) the y axis represents the mean blood pressure of each neighbourhood computed by aggregating the individual values in A and B. Figures C and D represent a classic ecological analysis.

Visual inspection of the ecological data (figs 4C and D) shows that the neighbourhood differences in mean blood pressure are of similar magnitude in both sets of data. However, considering the first ecological figure (C), the corresponding multilevel analysis shows that, the clustering of blood pressure values within neighbourhoods is very weak, even before taking into account the neighbourhood income status. In fact, differences between neighbourhoods only account for a very small portion of the total variance in blood pressure (that, the ICC is very low). In the second case (D), the corresponding multilevel analysis shows that the clustering of blood pressure values is very high in the empty model and still very important after taking the neighbourhood income into account as it can be seen in figure B2. In fact, in this extreme example, almost all variability between people is at the neighbourhood level (that, the ICC is very high). We see that the neighbourhood income variable explains a large part of the individual neighbourhood differences in SBP in case B, but only explains a very little share in case A.

Even if ecological observations C and D suggest that there are differences in blood pressure of similar magnitude, multilevel analyses A and B show that the neighbourhood differences over and above individual level differences are very small in C and very important in D.

The purpose of this extreme example is simply to show that an ecological analysis would provide similar information in cases A and B. The multilevel model, on the other hand, provides information on the size of clustering and shows that the two situations are completely different with respect to the



**Figure 4** This figure represents people living in neighbourhoods. We consider continuous SBP as the outcome variable, and the level of neighbourhood deprivation (measured as a standardised deprivation score) as the independent variable. We see that the magnitude of the regression coefficient (a measure of association) represented by the regression line is rather independent of the size of the intraclass correlation ICC (that is, the clustering of the outcome within neighbourhoods). The regression coefficient for the effect of the neighbourhood deprivation is about 1.2 mm Hg in A2, B2, C, and D; however, the size of the intraclass correlation is very low in example A1 and A2 (upper left) and very high in example B1 and B2 (upper right). In the first case, the clustering is very low and we could conclude that the context has no importance for understanding individual variation in blood pressure. In the second case, the clustering of individual SBP levels is high and the geographical environment has a very great influence on the individual outcome. Despite this, the sizes of the regression coefficients are similar. The regression coefficient reveals nothing of the magnitude of the remaining variability existing within neighbourhoods after adjustment for neighbourhood income, which is much higher in B2 than in A2 (a similar figure has been presented on a previous publication).<sup>5</sup>

relevance of considering neighbourhoods in seeking to understand individual health disparities.

### Including the concept of clustering when interpreting the effect of neighbourhood level variables

When studying measures of association (for example, regression coefficients), we aim to obtain information on possible causal relations between neighbourhood

characteristics and health. Figure 4 presents a clear association between neighbourhood standardised mean income and blood pressure with a regression coefficient roughly equal to 1.2 mm Hg. This association is similar in A2, B2, C, and D, but the regression coefficient shows nothing of the magnitude of clustering within neighbourhoods that remains after adjustment for the neighbourhood income, which is much larger in B2 than in A2.

Therefore, on the one hand, an association of a given magnitude between a neighbourhood variable and health may explain a large share or alternatively very little of the total individual level variations. And on the other hand, such association may coexist in the model with very small or alternatively very important residual intra-neighbourhood correlations. But the regression coefficient does not provide such information.

The existence of a small ICC in the empty model does not necessarily imply that it is not worth investigating the effects of neighbourhood variables on individual level outcomes.<sup>17</sup> At the same time, however, the existence of an ICC close to 0 does suggest that the context defined by the neighbourhoods is not that important in understanding individual health disparities.

It could be the case that a neighbourhood characteristic, even if significantly associated with the outcome, has almost no importance if it only explains a very small portion of neighbourhood variance. We can also imagine a case in which the neighbourhood variable would explain a large portion of a neighbourhood variance that is itself very small. In that case, explaining a large portion of neighbourhood variance will not make the neighbourhood variable important for public health, as a lot of very little is still very little.<sup>18</sup>

By neglecting information on clustering, ecological analyses—as well as the exclusive consideration of measures of association between neighbourhood characteristics and individual health—may lead to the stigmatisation of certain neighbourhoods. Indeed, in ignoring the variability within neighbourhoods, we may assume that the neighbourhood effect applies directly to all the people who live in it, which may lead to inappropriate area stereotypes such as “healthy” or “unhealthy” neighbourhoods. Stigmatisation may lead to neighbourhood blame and discrimination. If inaccurate perceptions of contextual effects are applied to health policy, such policies may be inefficient in reducing inequalities.<sup>19</sup> Therefore, taking into account the magnitude of clustering is especially relevant in the current research on neighbourhood and health. Unfortunately, many published articles do not inform on neighbourhood variance and clustering or simply avoid interpreting it.

Using measures of clustering of individual health status within neighbourhoods, we could evaluate the relative importance of the neighbourhood level for various kinds of outcomes, and may promote neighbourhood level interventions for those health outcomes that are in a greater extent than others determined by the neighbourhood. These considerations are important when attempting to determine the efficacy of focusing intervention on places instead than on people.<sup>7</sup> For example, if an intervention were to be focused on a given selection of low income neighbourhoods when in fact neighbourhood variation represents only a very small part of the total variation in the health outcome, then many people at risk would be ignored simply because they do not reside in low income neighbourhoods. When the clustering of individual health status within neighbourhoods is small, focusing intervention on specific places may be a rather inefficient strategy.<sup>7</sup>

Combining measures of variation and measures of association is therefore useful in assessing the importance of the neighbourhood environment and its characteristics when seeking an understanding of individual health inequalities.

In this article neighbourhoods were defined as geographical areas. However, multilevel modelling lends itself to a variety of other meanings of community and context including groups defined by cultural rather than by strict geographical administrative areas.<sup>20</sup>

In this and two previous articles<sup>12–13</sup> we have presented a number of concepts that have not been commonly discussed

in multilevel epidemiology despite their importance. Our essay contributes to an understanding of the connection between MLRA and social epidemiological concepts and of the relevance of these concepts for public health, with a special focus on the notion of contextual phenomenon.

#### Authors' affiliations

**J Merlo, B Chaix, L Råstam**, Department of Community Medicine (Preventive Medicine), Malmö University Hospital, Lund University, Malmö, Sweden

**B Chaix**, Research Team on the Social Determinants of Health and Healthcare, National Institute of Health and Medical Research, Paris, France

**M Yang**, Institute of Community Health Sciences, Queen Mary University of London, London, UK

**J Lynch**, Department of Epidemiology, Center for Social Epidemiology and Population Health, University of Michigan, Ann Arbor, USA

Funding: this study was supported by grants from FAS (Swedish Council for Working Life and Social Research) for the projects “Development and application of multilevel analysis in pharmacoepidemiology and social medicine” (principal investigator Juan Merlo, no 2002-054) and “Socioeconomic disparities in cardiovascular diseases—a longitudinal multilevel analysis” (principal investigator Juan Merlo, no 2003-0580).

Competing interests: none declared.

#### REFERENCES

- 1 Durkheim E. *The rules of sociological method*. 8th ed. New York: Free Press of Glencoe, 1964.
- 2 Rose GA. *The strategy of preventive medicine*. Oxford: Oxford University Press, 1992.
- 3 Snow J. *Snow on cholera being a reprint of two papers by John Snow, MD, together with a biographical memoir by B W Richardson, MD, and an introduction by Wade Hampton Frost, MD*. New York: The Commonwealth Fund, 1932.
- 4 Macintyre S, Ellaway A. Ecological approaches: rediscovering the role of the physical and social environment. In: Berkman L, Kawachi I, eds. *Social epidemiology*. New York: Oxford University Press, 2000:332–48.
- 5 Merlo J, Östergren 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.
- 6 Petronis KR, Anthony JC. A different kind of contextual effect: geographical clustering of cocaine incidence in the USA. *J Epidemiol Community Health* 2003;**57**:893–900.
- 7 Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *J Epidemiol Community Health* 2003;**57**:550–2.
- 8 Goldstein H. *Multilevel statistical models*. 3rd ed. London: Hodder Arnold, 2003.
- 9 Snijders TAB, Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. 1st ed. Thousand Oaks, CA: Sage, 1999.
- 10 Raudenbush SW, Bryk AS. *Hierarchical linear models: applications and data analysis methods*. 2nd ed. Thousand Oaks CA: Sage, 2002.
- 11 Subramanian SV, Jones K, Duncan K. Multilevel methods for public health research. In: Berkman L, Kawachi I, eds. *Neighbourhoods and health*. Oxford: Oxford University Press, 2003:65–111.
- 12 Merlo J, Chaix B, Yang M, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health* 2005;**59**:443–500.
- 13 Merlo J, Yang M, Chaix B, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *J Epidemiol Community Health* 2005;**59**:729–36.
- 14 Merlo J, Asplund K, Lynch J, et al. Population effects on individual systolic blood pressure: a multilevel analysis of the World Health Organisation MONICA project. *Am J Epidemiol* 2004;**159**:1168–79.
- 15 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. (<http://www.mlwin.com/beta/chapter07.pdf>).
- 16 Blakely TA, Woodward AJ. Ecological effects in multi-level studies. *J Epidemiol Community Health* 2000;**54**:367–74.
- 17 Duncan GJ, Raudenbush SW. Assessing the effects of context in studies of child and youth development. *Educ Psychol* 1999;**34**:29–41.
- 18 Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *J Educ Behav Statist* 1998;**24**:333–5.
- 19 Curtis S, Jones IR. Is there a place for geography in the analysis of health inequality? *Sociology of Health and Illness* 1998;**20**:645–72.
- 20 Beckman A, Merlo J, Lynch JW, et al. Country of birth, socioeconomic position, and healthcare expenditure: a multilevel analysis of Malmö, Sweden. *J Epidemiol Community Health* 2004;**58**:145–9.



# Speaker's corner

## The question is *not*: "Is race or class more important?"

Discussions about links between race and class are not new, but a recent resurfacing in the USA of the flawed question about which is more important leads me to choose this topic. While race may draw more attention here than in some other nations, the hideous genocide in the Sudan and ugly racial/ethnic divisions in many other places as well suggest that this discussion may have some wider relevance.

The issue of socioeconomic influences on health was marginalised in the USA until the Cold War ended. Since the early 1990s, there has been an opening for academic research on the mechanisms explaining class disparities in health. However, the opening in research has not been accompanied by major policy initiatives applying what is already known about how socioeconomic factors, particularly those experienced in childhood, constrain people's opportunities to be healthy throughout the life course.

Racism also has been a taboo topic. Awareness has been somewhat increased by recent reports from prestigious agencies, reviewing a large body of evidence showing systematic bias (probably largely unconscious) along racial/ethnic lines in health care delivery. Many public and private initiatives have been launched over the past decade, addressing racial/ethnic disparities in health care, most focusing on "cultural competence" by care providers as the remedy, sidestepping consideration of both structural racism and class.

Recently, the US Congress discussed replacing initiatives on racial/ethnic disparities with targeting the economically disadvantaged, regardless of race. Although the acknowledgement of economic disadvantage is welcome, this would

be an unfortunate move. It would take race and racism off the agenda, subsuming it under class. It also would play off people of colour against low income white people because resources would not be increased along with the widened scope. One recurring theme among proponents of this direction is that class is "more important" than race.

Weighing race and class against each other is based on a false construct. Given the history of racism, even if all interpersonal racism were eliminated, the institutional structures that shape people's opportunities in life from early childhood—for example, by determining their parents' and their own educational and employment opportunities—are so powerful that we would continue to see huge racial/ethnic differences in many health outcomes.

Furthermore, while we cannot understand how racism operates without understanding how it systematically sorts people into different socioeconomic positions, giving them different opportunities, race and racism cannot be reduced to class and class prejudice. The widened African American/European American disparity in birth outcomes among women of higher socioeconomic status/position illustrates this, as does the paradox of worse birth outcomes among US born Latinas and US born African Americans compared with their often economically worse off immigrant counterparts.

An exclusive focus on class ignores or downplays the role of racism. The question is not which is more important, race or class, but how do we address both, including their complex interactions, with research and effective action.

**Paula Braveman**

Center on Social Disparities in Health, University of California, San Francisco, San Francisco, USA; Braveman@fcm.ucsf.edu