



Individual and collective bodies: using measures of variance and association in contextual epidemiology

Juan Merlo, Henrik Ohlsson, Kristian F Lynch, Basile Chaix, S V Subramanian

► To cite this version:

Juan Merlo, Henrik Ohlsson, Kristian F Lynch, Basile Chaix, S V Subramanian. Individual and collective bodies: using measures of variance and association in contextual epidemiology. *Journal of Epidemiology and Community Health*, 2009, 63 (12), pp.1043-n/a. 10.1136/jech.2009.088310 . hal-00484148

HAL Id: hal-00484148

<https://hal.science/hal-00484148v1>

Submitted on 18 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Individual and collective bodies: using measures of variance and association in contextual epidemiology

Juan Merlo¹, Henrik Ohlsson¹, Kristian F Lynch^{1,2}, Basile Chaix^{1,3}, SV Subramanian⁴

¹ Unit for Social Epidemiology, Faculty of Medicine, Lund University, Malmö, Sweden

² Unit for Diabetes and Celiac Disease, Department of Clinical Sciences, Faculty of Medicine, Lund University, Malmö, Sweden

³ UMR-S 707 Inserm, Université Pierre et Marie Curie, Paris 6, Faculté de Médecine Saint-Antoine, Paris, France

⁴ Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, MA, USA

Corresponding author:

Prof. Juan Merlo MD, PhD
Unit for Social Epidemiology, Faculty of Medicine, Lund University
CRC, UMAS, Entrance 72, house 28
SE-20502 Malmö
Sweden
E-mail: juan.merlo@med.lu.se
Tel: +46 40 301329

Keywords: social medicine, multilevel analysis, analysis of variance, causality, social environment

Word count: about 3680

ABSTRACT

Background

Social epidemiology investigates both individuals and their collectives. While the limits that define the individual bodies are very apparent, the collective body's geographical or cultural limits (e.g., "neighbourhood") are more difficult to discern. Also, epidemiologists normally investigate causation as changes in group means. However, many variables of interest in epidemiology may cause a change in the variance of the distribution of the dependent variable. In spite of that, variance is normally considered a measure of uncertainty or a nuisance rather than a source of substantive information. This reasoning is also true in many multilevel investigations, whereas understanding the distribution of variance across levels should be fundamental. This means-centric reductionism is mostly concerned with risk factors and creates a paradoxical situation, since social medicine is not only interested in increasing the (mean) health of the population, but also in understanding and decreasing inappropriate health and health care inequalities (variance).

Methods

Critical essay and literature review.

Results

The present essay promotes (a) the application of measures of variance and clustering to evaluate the boundaries one uses in defining collective levels of analysis (e.g., neighbourhoods), (b) the combined use of measures of variance and means-centric measures of association, and (c) the investigation of causes of health variation (variance-altering causation).

Conclusions

Both measures of variance and means-centric measures of association need to be included when performing contextual analyses. The variance approach, a new aspect of contextual analysis that cannot be interpreted in means-centric terms, allows us to expand our perspectives.

People are simultaneously social and biological organisms and therefore social epidemiology distinguishes itself from general epidemiology by its inherent multilevel approach that aims to investigate both individuals and their collectives all together.[1] While the limits that define the individual biological bodies are very apparent, the collective body's geographical or cultural limits (e.g., "neighbourhood") are more difficult to discern. Also, it is normal in epidemiology to investigate causation as changes in group means **even though many variables of interest may cause a change in the variance of the distribution of the dependent variable and not cause a change in the mean. So far here has been little** interest in understanding changes in terms of the variance that underlies averages. [2-6] Variance is often considered a measure of uncertainty or a troublesome entity, rather than a source of substantive information. Paradoxically, this restrictive approach is the norm in many multilevel investigations, [7] whereas understanding the distribution of variance across levels should be the sine qua non of any solid analysis. [8-16] One should always remember that the goal of social medicine is not only to increase the (mean) health of the population, but to decrease health **and health care** inequalities (variance).

The present **essay** questions the usual means-centric approach and emphasises the need to deliberately investigate the heterogeneity that underlies averages. We propose, (a) the application of measures of variance and clustering to evaluate the boundaries that are to be used in defining collective levels of analysis (e.g. neighbourhoods). Further, we illustrate how a better understanding of contextual effects may be had by also (b) combining measures of variance with means-centric measures of association, and by (c) investigating the causes of health variation (variance-altering causation), rather than only considering changes in averages.

A. MEASURING VARIANCE TO EVALUATE THE BOUNDARIES DEFINING COLLECTIVE LEVEL OF ANALYSIS

Individual bodies, collective bodies, and Frankenstein

Researchers traditionally, investigate average characteristics of individuals (e.g., blood pressure) or of areas (e.g., social cohesion), but seldom question the “boundaries” that define the units of analysis. These boundaries are often accepted a priori. At the individual level, the limits that define the human body are very apparent, and so is the intra-individual correlation of the individual parts. Without question there is a *generalised individual effect* that maintains a sophisticated homeostasis among an array of cellular and physiological processes (e.g., blood pressure level) within the physical boundaries of our skin. As a result, when performing multiple measurements of individual variables in a group of subjects, a great proportion of the variance between measurements is at the individual level. From an epidemiological perspective this intra-individual correlation is a nuisance that, because of statistical reasons, needs be accounted for when analysing such things as repeated blood pressure measurements, diseases of the eyes or teeth, or multiple bone fractures. However, since the boundaries of human bodies are so well delineated, we can take them for granted. Instead, our research mainly focuses on whether exposure to a variable (i.e., antihypertensive medication) has an effect on, or changes, the mean of the distribution of another variable (i.e., lowering of blood pressure).

Even since the days of Durkheim [17] we know that when individuals adhere to each other and form a coherent community, a relational [18] and collective effect emerges that becomes more than the sum of each individual action. On some occasions this collective effect may have arisen because of sharing of common geographic environments, [19] and not necessarily as a result of a voluntary individual decision to form a social group. In any case, this *general collective effect* is to some extent analogous to the general individual effect

mentioned above. Like multiple measurements, individuals within a collective are themselves more similar to one another than they are to individuals outside of their group because insiders share common social boundaries and contextual influences. [10] As in the case of the individual body, the collective body maintains a social homeostasis that balances numerous social processes. However, lacking a covering skin such as human bodies possess, the collective body's geographical or cultural limits (e.g., "neighbourhood") are more difficult to discern. A further source of complexity exists: unlike the parts of a human body, [20] individuals often belong to several collectives, [21,22] both in a cross-sectional and in a life course perspective, [23]. Sometimes, however, collective boundaries are relatively easy to recognise, as is the case of schools or health care centres when investigating pupil outcomes [24] or medical treatments. [25]

While the existence of contextual effects on individual health is rather obvious, [26] the validity of many administrative boundaries for defining collective bodies [27,28] requires more investigation. [18] In fact, we cannot exclude the possibility many of the "neighbourhoods" we study are – like contextual Frankensteins – constructed by assembling parts from different collective bodies. Paraphrasing Duncan, Jones, & Moon: "the hierarchical definition of the levels could be criticized as an inappropriately formalistic and mechanistic attempt to capture the cultural geography of lifestyle". [29], p. 732

Can we identify appropriate collective boundaries for investigating contextual effects?

While a number of collective bodies (e.g., neighbourhoods) are assumed to be exclusively delimited by geographical boundaries, other cultural or relational criteria (either alone or combined with geographical information) may be more appropriate for identifying collective bodies. [18,30,31] In any case, if such a collective body exists, one might anticipate finding a correlation in the health of the individuals within it. [10] Moreover, the closer our boundary

definition corresponds to the boundaries of the hypothesized collective body, the greater one may expect the clustering of the health outcome of interest to be. [10,32] Based on this assumption we can try to identify collective boundaries in one of two ways: either by (a) scanning a geographical surface to identify clusters of the outcome (e.g., specific diseases), or by (b) using a priori defined socio-geographical areas and quantifying the observed clustering of disease outcomes within these areas.

In the first instance above, cluster recognition analysis [33,34] may help us to identify geographical clusters of the outcome of interest (e.g., the incidence of cholera). Once the relational [18] boundaries that include the clusters have been identified, we can investigate what makes them different from the rest of the geographical surface (e.g., the water supply) and thereby obtain relevant information for drawing causal hypotheses and planning public health interventions. This idea is close to John Snow's approach in his seminal work on cholera. [35] In the second scenario (b) the most common procedure is to employ geographical divisions found in administrative databases (e.g., census tracts). Alternatively, such areas may be combined on the basis of certain characteristics [31,36,37] or by using Geographic Information Systems (GIS) techniques. [34]

Measures of variance and clustering (e.g., intraclass correlation, [15,38] median odds ratio, [9,39,40] or pairwise odds ratio [32]) allows us to identify the scale (e.g., local neighbourhoods, parishes, or municipalities) [41,42] on which contextual influences operate with different health outcomes.[43] For example, in 1999 Boyle & Willms performed an unconventional multilevel analysis exclusively based on measures of intra-class correlation and places defined by administrative boundaries. [44] They observed that place effects were generally small and were influenced by both the size of the geographical area used to define place and the health indicator selected for study. The authors questioned the usefulness of carrying out health needs assessment surveys within large administrative areas, and casted

doubt on the utility of these geographic boundaries for studying place effects. Boyle & Willms based their conclusions not on means-centric measures of association, but on the analysis of variation, and they raised questions about the context as a whole. Their work, has resulted in some criticism (see, for example, Blakely [45] p. 373).

B. COMBINING MEASURES OF VARIANCE AND MEAN-CENTRIC MEASURES OF ASSOCIATION FOR A BETTER UNDERSTANDING OF CONTEXTUAL EFFECTS

Small variance but conclusive associations

It may appear paradoxical that a conclusive (“significant”) association between a contextual variable and an individual outcome can be detected alongside a very small fraction of overall variance in the outcome at the contextual (e.g., neighbourhood) level. [46] Thus, means-centric measures of association indicate the existence of contextual effects, while measures of variance suggest the opposite. In an attempt to justify this paradox [47,48], it has been commented that **standardized mean differences (d) between intervention and control neighbourhoods, that program evaluators commonly view as medium (d = 0.4) or even large (d = 0.6) translate into ‘small’ intra-neighbourhood correlations of 4% and 8% respectively.** Several criticisms can be raised to this reasoning, with the most fundamental being there is no need for such a justification. As our essay argues, such paradox does not exist but, rather, the question is to distinguish between the different information provided by measures of association and by measures of variance. Furthermore, it has been **previously** revealed that **standardised coefficients [47] are inappropriate measures of effect as they can be “confounded” by the variance of the specific setting where the study is performed (see Greenland [49] and Cummings[50] for an extended explanation).** Standardized effects are

actually distorted measures of association and also hide the information contained in the variance being standardized..

In addition, while intraclass correlations ranging from 4% - 8 % may be considered small by some [47], they may be highly relevant for others particularly when compared with the intraclass correlations of many “neighbourhood” outcomes relevant for social epidemiology. [51]

The fundamental difference between measures of variation and means-centric measures of association becomes clear if one understands that the interpretation of variance is often temporo-spatially constrained and that for every individual outcome there may be a pattern of variance produced by different environmental conditions. [52] Thus, in seeking useful information for planning public health interventions, measures of variance and clustering from a specific context provide insight into the areas being investigated during the period of the study. By contrast, measures of association intend to provide causal information that can be generalised and applied to contexts beyond the one where the study was performed.

Further illustration of this idea using a continuous contextual variable is given in the figure 1 and its accompanying legend. See also a previous study published elsewhere in this Journal.[11])

-[Figure 1 about here]

We can even clarify this apparent contradiction in an equation that can be calculated from a two level logistic regression model. As an example, imagine individuals are nested within neighbourhoods and the aim is to analyze the relation between neighbourhood deprivation (X) and presence of hypertension. Without loss of generality, assume the

contextual variable is centred on the mean. It can be shown that the total neighbourhood variance σ^2_T (as originally estimated in the simplest “empty” model with only individual nested within neighbourhoods) is a function of the regression coefficient of the contextual variable (β^2_X) the variance of this contextual variable (σ^2_X), and of the residual variance (σ^2_u) once the contextual variable is included in the model. Therefore it is possible to find a similar β_X with very different scenarios of variance.

$$\sigma^2_T = \beta^2_X \cdot \sigma^2_X + \sigma^2_u \quad \text{Equation 1}$$

Failure to distinguish between the two types of measures explained above, and specially, interpreting means-centric measures of association as if there were measures of variance may lead to inappropriate conclusions, a situation that is unfortunately rather common in many multilevel analyses performed today. As one example, in 2001 Diez-Roux et al, [53] performed a state of the art and thoroughly conducted multilevel analysis of the relationship between characteristics of neighbourhoods and the incidence of coronary heart disease. Possibly, since the samples of individuals within blocks in this Atherosclerosis Risk in Communities Study (ARIC) were very small, the authors did not estimate neighbourhood variance. They did, however, assume the existence of intra-neighbourhood correlation in the outcome, which was considered a statistical nuisance and overcome by adjusting standard errors for clustering using the statistical software SUDAAN. The size of this conjectural intra-neighbourhood correlation was never reported. The study drew two main conclusions: (1) “Neighbourhood characteristics are related to the incidence of coronary heart diseases” and (2) “Strategies for disease prevention may need to combine person-centered approaches with approaches aimed at changing residential environments.” While these conclusion are of clear academic interest for understanding the contextual causes of coronary heart disease, they are also vague and may even be misleading. An approach aimed at changing residential environments may be effective in the ARIC context, but since we do not have information on

the intra-neighbourhood correlation (or similar measures of variance), we cannot be certain. For example, although an analogous multilevel study in Malmö, Sweden, [54] demonstrated a clear association between the socioeconomic characteristics of the urban areas and individual blood pressure (in agreement with the ARIC study) [53], differences between these same areas explained less than 1% of the individual variance in blood pressure. Consequently, the study concluded that, with regard to Malmö, an intervention focused on urban areas with a higher mean level of deprivation would be ineffective [4,11]

We should not avoid a contextual analysis because the area variance or the intra-class correlation is very small, since means-centric associations between contextual variables and individual health may still be detected. On the other hand, neither can one recommend a contextual public health intervention based on a “significant” means-centric association if the clustering of individual health within areas is unknown or very low. Means-centric measures of association do not provide sufficient information for deciding to launch public health interventions **at some specific areas but not at other**. In fact, if clustering is small, a public health intervention would be ineffective, even where a contextual variable is associated with the individual outcome and serves to explain 100% of the area variance. As Singer has stated “You can explain a large amount of very little” [55], p. 332. A different scenario is also possible in which we observe a clear association between the contextual variable and the individual outcome side by side with a very large residual area variance. In this case, if the contextual variable does not go very far toward explaining the original area variance, a public health intervention aimed at changing this contextual variable would not be very ineffective.

As discussed earlier, measures of variance and clustering are useful for identifying boundary limits of the “collective body” that we assume influence the outcome under study. In this section we sought to demonstrate that such collective boundaries (i.e., “the geographic scale”) cannot be properly identified solely by means-centric measures of

association. For this purpose, means-centric measures of association must be combined with measures of variance and clustering.

New analytical approaches – solving the paradox

In an earlier publication, [9] we have given empirical examples of the two scenarios depicted above, and have also sought to apply a pair of useful measures proposed by Larsen et al [9,40]: the median odds ratio (MOR) and the interval odds ratio (IOR) (see the Tutorial published elsewhere in the Journal for a extended explanation of these measures)[39]. The MOR quantifies neighbourhood variance on the odds ratio scale, and the IOR incorporates both the means-centric effect (i.e., odds ratio) and neighbourhood variance in one interval, allowing for a more detailed description of the means-centric effect.

In our previous study, [9] we found the administrative neighbourhoods of Malmö were very suitable for identifying the “collective body” that conditions certain individual behaviours, such as choosing to visit a private rather than a public physician. For this condition, the area variance (SE) with adjustment for age and individual education was 1.815 (0.278) and the correspondent MOR = 3.61. However, the same neighbourhoods seemed inappropriate for identifying the “collective body” that conditioned hospitalisation for ischemic heart disease, since the corresponding variance (SE) was only 0.028 (0.025) and the MOR = 1.17. In the same study, the socioeconomic characteristics of the neighbourhood appraised by aggregated educational achievement (low vs. high) were, however, associated with both outcomes. Nevertheless, in spite of this observed association (and disregarding concerns about counter-factuality) we concluded that a possible public health intervention directed to specific neighbourhoods would be ineffective in either instance. In the case of hospitalisation for ischemic heart disease the inefficiency would depend on the very low neighbourhood variance. Finding such a low area variance in cross-sectional studies is rather common for chronic diseases which, like arteriosclerosis, develop over a whole life-course

and have little to do with the place where an individual may be currently residing. When it comes to change individual's choice of physician, a neighbourhood intervention would similarly be ineffective because of the large variance that remains unexplained after including the variable "neighbourhood level of education" in the model. The latter expressed itself in a very broad IOR = 0.28–27.3. There is little doubt that the current neighbourhood context influences individual behaviour much more than chronic disease, but "neighbourhood educational level" – in spite of being associated with the choice of physician – does not explain very much in the city of Malmö.

We state that combining variance-based measures with means-centric measures of association provides useful and complementary information on contextual effects. These considerations may be relevant when attempting to determine the efficacy of focusing intervention on places rather than on people. For example, imagine that a City Council has been informed that average blood pressure is higher in deprived neighbourhoods than in wealthy neighbourhoods. As a consequence, decision makers are considering the allocation of resources in the most deprived neighbourhoods for the creation of new health care centres specialized on blood pressure control. However, if the neighbourhood variation represents only a very small part of the total individual variation in blood pressure, then many people with high blood pressure would be ignored simply because they reside in wealthy neighbourhoods. When the clustering of individual health status within neighbourhoods is small, focusing intervention on specific places may be a rather inefficient strategy. [4,11] Using the words of Clarke: "without knowledge of the random components, the interpretation of area-level fixed effects parameters becomes *decontextualized*".[37] p. 315.

C. INVESTIGATING THE CAUSES OF HEALTH VARIATION

Epidemiologists commonly understand causation in terms of group means, so the statement - “X causes Y” is taken to imply that, *ceteris paribus* (see references [14,48,56,57] for a discussion on this aspect), an increase in the value of X changes the mean of the distribution of Y. However, many independent variables of interest in epidemiological studies may cause a change in the variance (not the mean) of the distribution of the dependent variable [2-6] (see figure 2).

[Figure 2 about here]

The distinction between the variance altering and means-centric altering approaches is still not widely observed in social epidemiology; most researchers only discuss classic means-centric measures of association. This means-centric reductionism goes hand by hand with an epidemiology mostly concerned with risk factors [58] and drug safety and creates a paradoxical situation, since social medicine is not only interested in increasing the (mean) health of the population, but also in decreasing health inequalities (variance). Likewise, it is of major relevance to understand and prevent inappropriate health care variation, as it leads to inefficient resource utilization. Modelling variance itself as a dependent variable may provide useful information on health inequalities and suggest a different kind of contextual effect. [6,13,25,59]

When investigating such variance-altering causation in contextual analysis, a fundamental independent variable is the definition of boundaries that we use to operationalise collective bodies (as we have attempted to show in the first section above). For example, we include ‘neighbourhood’ as a random term in multilevel regression analyses. We can consider the boundaries that define a specific level of analysis as an independent variable in an equation that models variance. In this way we can investigate (potential) causation where the

chosen boundaries – assuming they delimit a true collective body – “cause” a certain pattern of individual differences/similarity. This **has been the** approach adopted, for example, by Boyle & Willms [44], **Reijneveld [60], Petronis & Anthony [32]** and by ourselves [6,25,46,54] as well as recently **by Uthman, Moradi & Lawoko [61] and by** Naess et al [62]. **This last work** explores area variation across a life-course as a way of elucidating potential (causal) influence of area on mortality.

Modelling individual and area variance may yield valuable information on how contextual factors shape health inequalities for different individuals. In a previous study based on the MONICA project [6] we investigated contextual effects on individual systolic blood pressure, and modelled variance as a function of antihypertensive medication use and body mass index. Among other results we found that contextual effects were particularly strong in overweight women on antihypertensive **medication. Actually, around 20% of the individual differences in blood pressure were conditioned by the MONICA population where these women were included. This contextual phenomenon** possibly reflects **disparities in the effectiveness of antihypertensive treatment among different** national health care systems (see figure 3 in reference [6]).

In a recent study [25] we have explicitly investigated variance altering causes and presented a conceptual illustration showing that a change in the characteristics of a context (e.g., the implementation of a decentralized health care budget) not only changes the mean of the distribution of the variable studied (i.e., increasing compliance with prescription guidelines), but alters the variance between the collective units as well (i.e., decreasing inequality between health care centres). Downs & Rocke [2] and Braumoeller [5] also provide illustrative examples in their work.

CONCLUSIONS

Both measures of variance and means-centric measures of association need to be included when performing contextual analyses. However, more research is needed to (a) identify appropriate boundaries for collective bodies like neighbourhoods, (b) develop statistical methods that facilitate the use of measures of variance in social epidemiology, (c) identify variance-altering causes and their mechanisms, and (d) comprehend the relationship among the degree of clustering of individual health within administrative areas, the size of means-centric measures of association, and the possible efficiency of public health interventions.

Seeking causal explanations in social epidemiology is a challenge in itself,[63] but focusing on the (causal) circumstances that condition variance reveals a neglected theoretical dimension for understanding health disparities in social epidemiology. The variance approach, a new aspect of contextual analysis that cannot be interpreted in means-centric terms, allows us to expand our perspectives.

ACKNOWLEDGEMENTS

This study was supported by grants from the Swedish Research Council (PI: Juan Merlo, Dnr: 2004-6155); the Centre for Economic Demography, Lund University; the Swedish Council for Working Life and Social Research (Juan Merlo, Dnr: 2007-1772); an ALF Government Research Grant (Juan Merlo, Dnr: M: B 39 977); and by a National Institutes of Health Career Development Award (S.V. Subramanian NHLBI K25 HL081275). We want express our gratitude to Prof. Fiona Steele (The Centre for Multilevel Modelling, University of Bristol, UK) for her comments on the last version of this article.

Contributorship Statement

JM had the original idea of this study and wrote the manuscript. HO, KFL, BC and SS have actively contributed to the discussion and interpretation of the ideas presented in this essay. All authors have critically revised this article for important intellectual content. All the authors have read and approved the final version to be published.

Licence Statement

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in JECH editions and any other BMJPGL products to exploit all subsidiary rights, as set out in our licence (<http://jech.bmj.com/ifora/licence.pdf>).

REFERENCES

1. Krieger N. A glossary for social epidemiology. *J Epidemiol Community Health* 2001;**55**:693-700
2. Downs GW, Roche DM. Interpreting Heteroscedasticity. *American Journal of Political Science* 1979;**23**:816-828
3. Gould SJ. *Full House: The Spread of Excellence from Plato to Darwin* New York: Three Rivers Press; 1996
4. Merlo J. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *J Epidemiol Community Health* 2003;**57**:550-552
5. Braumoeller B. Explaining Variance; Or, Stuck in a Moment We Can't Get Out Of. . *Political Analysis* 2006;**14**:268-290
6. Merlo J, Asplund K, Lynch J, Rastam L, Dobson A. Population effects on individual systolic blood pressure: a multilevel analysis of the World Health Organization MONICA Project. *Am J Epidemiol* 2004;**159**:1168-1179
7. Roux AV. Next steps in understanding the multilevel determinants of health. *J Epidemiol Community Health* 2008;**62**:957-959
8. Riva M, Gauvin L, Barnett TA. Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. *J Epidemiol Community Health* 2007;**61**:853-861
9. Larsen K, Merlo J. Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. *Am J Epidemiol* 2005;**161**:81-88
10. Merlo J, Chaix B, Yang M, Lynch J, Rastam L. 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**:443-449
11. Merlo J, Chaix B, Yang M, Lynch J, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *J Epidemiol Community Health* 2005;**59**:1022-1028
12. Merlo J, Yang M, Chaix B, Lynch J, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *J Epidemiol Community Health* 2005;**59**:729-736
13. Subramanian SV, Jones K, Duncan C. Multilevel methods for public health research. In: Kawachi I, Berkman LF eds, *Neighborhoods and Health*. New York, NY: Oxford University Press; 2003:65-111
14. Subramanian SV. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Soc Sci Med* 2004;**58**:1961-1967
15. Goldstein H, Browne W, Rasbash J. Partitioning variation in generalised linear multilevel models. *Understanding Statistics* 2002;**1**:223-232
16. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med* 1998;**46**:97-117
17. Durkheim E. *The rules of sociological method*. 8th ed ed. New York: Free Press of Glencoe; 1964
18. Cummins S, Curtis S, Diez-Roux AV, Macintyre S. Understanding and representing 'place' in health research: A relational approach. *Soc Sci Med* 2007;**65**:1825-1838
19. Chaix B. Geographic Life Environments and Coronary Heart Disease: A Literature Review, Theoretical Contributions, Methodological Updates, and a Research Agenda. *Annu Rev Public Health* 2009;doi:10.1146/annurev.publhealth.031308.100158

20. McHugh K, Kiely EM, Spitz L. Imaging of conjoined twins. *Pediatr Radiol* 2006;**36**:899-910; quiz 1002-1003
21. Fielding A, Goldstein H. Cross-classified and multiple membership structures in multilevel models: an introduction and review. In, Research Report RR791: Department for University skills, University of Birmingham; 2006:<http://www.dcsf.gov.uk/research/data/uploadfiles/RR791.pdf>
22. Duncan C, Jones K, Moon G. Health-related behaviour in context: a multilevel modelling approach. *Soc Sci Med* 1996;**42**:817-830.
23. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol* 2002;**31**:285-293
24. Aitkin M, Longford N. Statistical modelling issues in school effectiveness. *J R Statist Soc A* 1986;**149**:1-43
25. Ohlsson H, Merlo J. Understanding the effects of a decentralized budget on physicians' compliance with guidelines for statin prescription--a multilevel methodological approach. *BMC Health Serv Res* 2007;**7**:68
26. Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med* 2002;**55**:125-139
27. Holt D, Steel DG, Tranmer M. Area homogeneity and the modifiable areal unit problem. *Geographical Systems* 1996;**3**:181-200
28. Diez Roux AV. Investigating neighborhood and area effects on health. *Am J Public Health* 2001;**91**:1783-1789
29. Duncan C, Jones K, Moon G. Do places matter? A multi-level analysis of regional variations in health-related behaviour in Britain. *Soc Sci Med* 1993;**37**:725-733
30. Beckman A, Merlo J, Lynch JW, Gerdtham UG, Lindstrom M, Lithman T. Country of birth, socioeconomic position, and healthcare expenditure: a multilevel analysis of Malmö, Sweden. *J Epidemiol Community Health* 2004;**58**:145-149
31. Knox PL, Pinch S. *Urban Social Geography: An Introduction*. 5th ed. Essex, England: Pearson Education Limited; 2006
32. 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
33. Estivill-Castro V, Lee I. Multi-level clustering and its visualization for exploratory spatial analysis. *GeoInformatica* 2002;**6**:123-152
34. Aldstadt J, Getis A. Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters. *Geographical Analysis* 2006;**38**:327-343
35. Snow J. Snow on cholera being a reprint of two papers by John Snow, M.D., together with a biographical memoir by B.W. Richardson, M.D., and an introduction by Wade Hampton Frost, M.D. New York: The Commonwealth Fund; 1936
36. Mujahid MS, Diez Roux AV, Morenoff JD, Raghunathan T. Assessing the measurement properties of neighborhood scales: from psychometrics to econometrics. *Am J Epidemiol* 2007;**165**:858-867
37. Clarke P, Wheaton B. Addressing Data Sparseness in Contextual Population Research: Using Cluster Analysis to Create Synthetic Neighborhoods. *Sociological Methods Research* 2007;**35**:311 - 351
38. Snijders TAB, Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. 1 ed. Thousand Oaks, California: SAGE Publications; 1999
39. Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health* 2006;**60**:290-297

40. Larsen K, Petersen JH, Budtz-Jorgensen E, Endahl L. Interpreting parameters in the logistic regression model with random effects. *Biometrics* 2000;**56**:909-914
41. Moellering H, Tobler WR. Geographical variances. *Geographical Analysis* 1972;**41**:34-50
42. Subramanian SV, Duncan C, Jones K. Multilevel perspectives on modeling census data. *Environment and Planning A* 2001;**33**:399-417
43. Murray DM, Hannan PJ, Jacobs DR, McGovern PJ, Schmid L, Baker WL, et al. Assessing intervention effects in the Minnesota Heart Health Program. *Am J Epidemiol* 1994;**139**:91-103
44. Boyle MH, Willms JD. Place effects for areas defined by administrative boundaries. *AmJEpidemiol* 1999;**149**:577-585
45. Blakely TA, Woodward AJ. Ecological effects in multi-level studies. *JEpidemiolCommunity Health* 2000;**54**:367-374
46. Lynch FK, Subramanian SV, Ohlsson H, Chaix B, Lernmark A, Merlo J. Context and disease when disease risk is low. The case of Type 1 Diabetes in Sweden. (working paper, available on request) 2008
47. Duncan GJ, Raudenbush SW. Assessing the effects of context in studies of child and youth development. *Educ Psychol* 1999;**34**:29-41
48. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc Sci Med* 2004;**58**:1953-1960
49. Greenland S, Schlesselman JJ, Criqui MH. The fallacy of employing standardized regression coefficients and correlations as measures of effect. *Am J Epidemiol* 1986;**123**:203-208
50. Cummings P. Meta-analysis based on standardized effects is unreliable. *Arch Pediatr Adolesc Med* 2004;**158**:595-597
51. Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess* 1999;**3**:iii-92
52. Lewontin RC. The analysis of variance and the analysis of causes. 1974. *Int J Epidemiol* 2006;**35**:520-525
53. Diez Roux AV, Merkin SS, Arnett D, Chambless L, Massing M, Nieto FJ, et al. Neighborhood of residence and incidence of coronary heart disease. *N Engl J Med* 2001;**345**:99-106
54. Merlo J, Ostergren PO, Hagberg O, Lindstrom M, Lindgren A, Melander A, et al. Diastolic blood pressure and area of residence: multilevel versus ecological analysis of social inequity. *J Epidemiol Community Health* 2001;**55**:791-798
55. Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growths model. *J Educ Behav Statist* 1998;**24**:333-335
56. Oakes JM. The (Mis)Estimation of Neighborhood Effects: Causal Inference for a Practicable Social Epidemiology. *Soc Sci Med* 2003
57. Merlo J, Chaix B. Neighbourhood effects and the real world beyond randomized community trials: a reply to Michael J Oakes. *Int J Epidemiol* 2006;**35**:1361-1363
58. Susser M. Does risk factor epidemiology put epidemiology at risk? Peering into the future. *J Epidemiol Community Health* 1998;**52**:608-611
59. Rasbash J, Browne W, Goldstein H, Yang M, Plewis I, Healy M, et al. Modelling the variance as a function of explanatory variables. In, *A user's guide to MLwiN* <http://multilevelioeacuk/download/usermanpdf>. version 2.1 ed: Multilevels Models Project, Institute of Education, University of London; 2000:77-88

60. Reijneveld SA, Verheij RA, de Bakker DH. The impact of area deprivation on differences in health: does the choice of the geographical classification matter? *J Epidemiol Community Health* 2000;**54**:306-313
61. Uthman OA, Moradi T, Lawoko S. The independent contribution of individual-, neighbourhood-, and country-level socioeconomic position on attitudes towards intimate partner violence against women in sub-Saharan Africa: A multilevel model of direct and moderating effects. *Soc Sci Med* 2009
62. Naess O, Claussen B, Smith GD, Leyland AH. Life course influence of residential area on cause-specific mortality. *J Epidemiol Community Health* 2008;**62**:29-34
63. Kaufman JS, Cooper RS. Seeking causal explanations in social epidemiology. *Am J Epidemiol* 1999;**150**:113-120

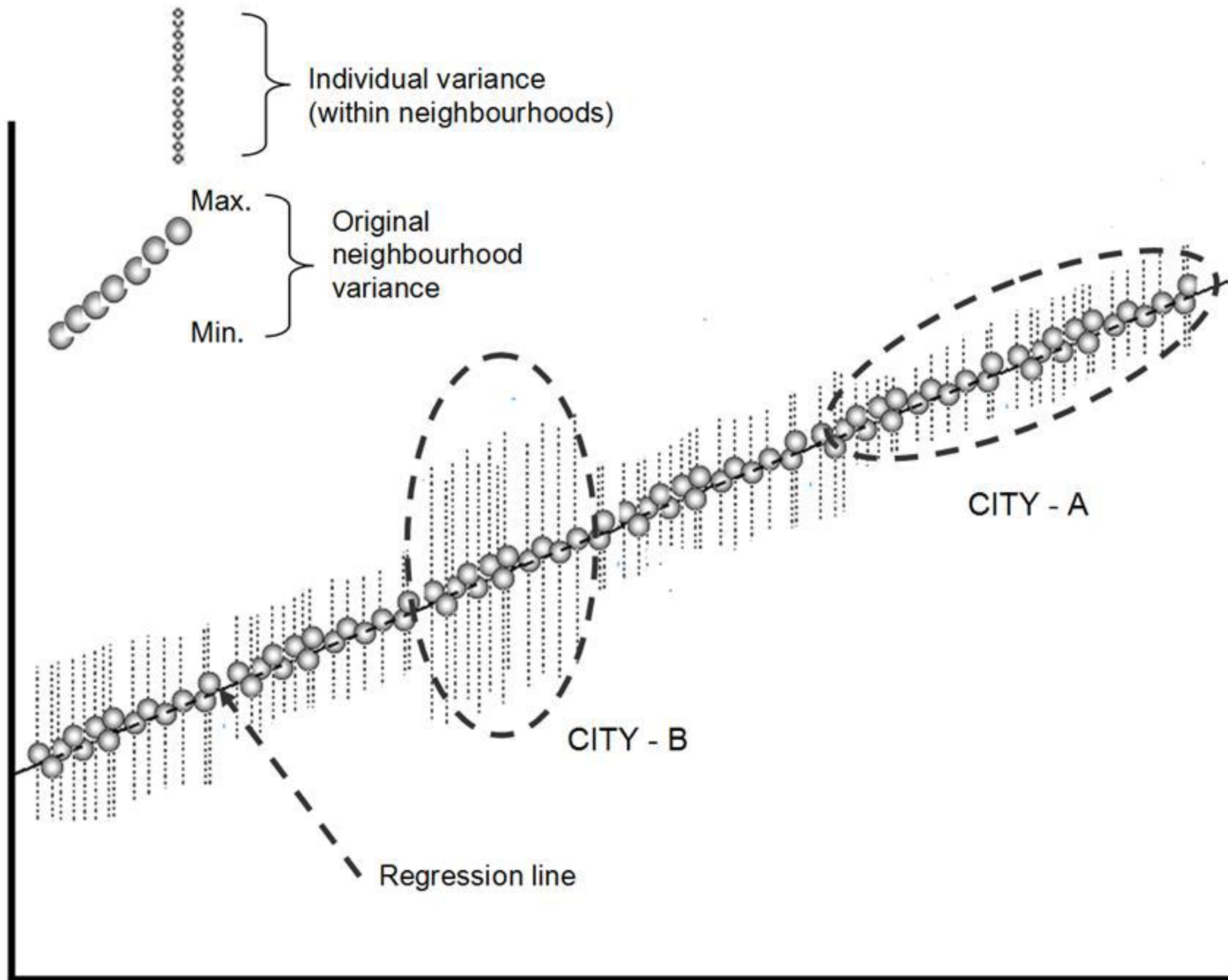
LEGEND TO THE FIGURES

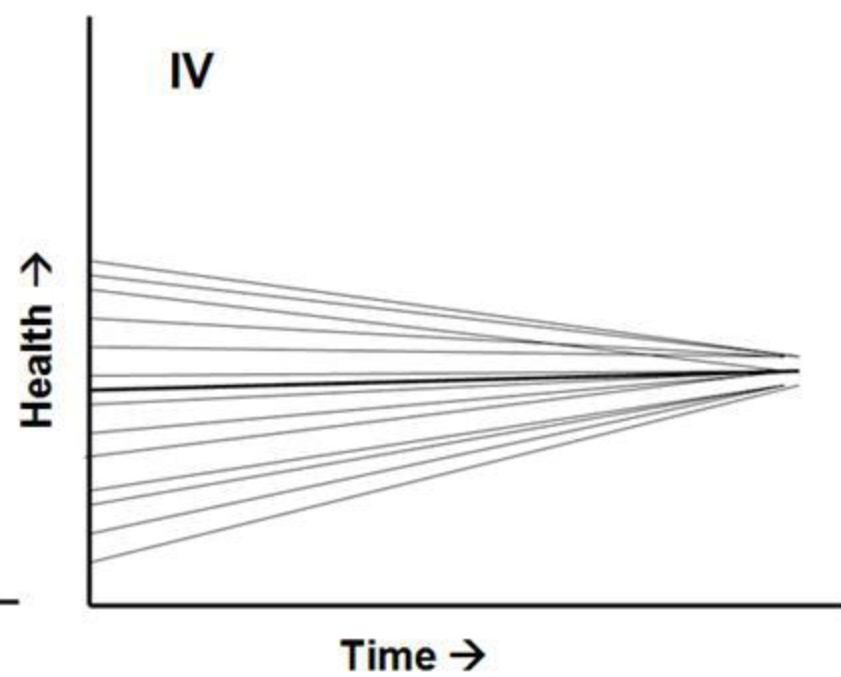
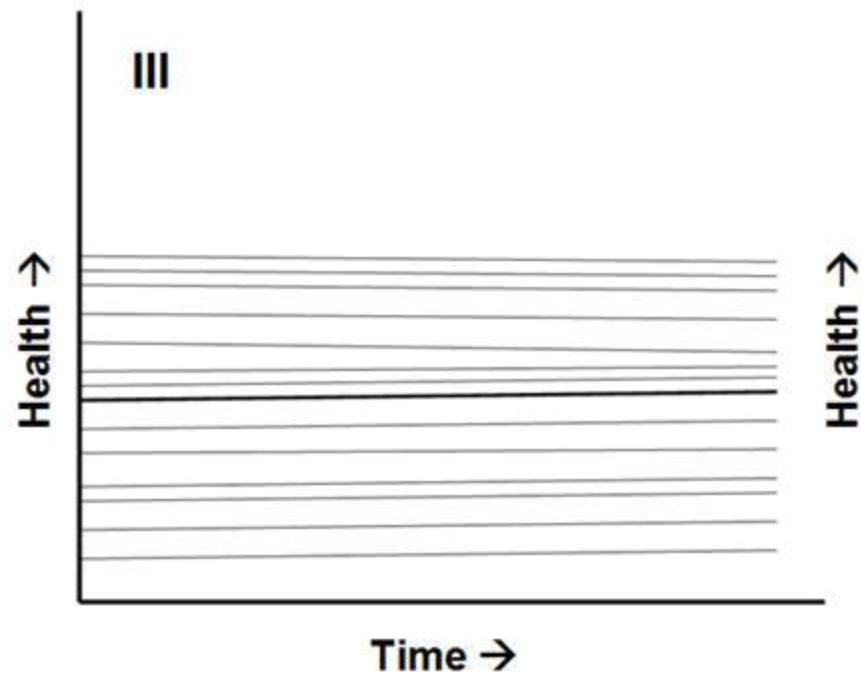
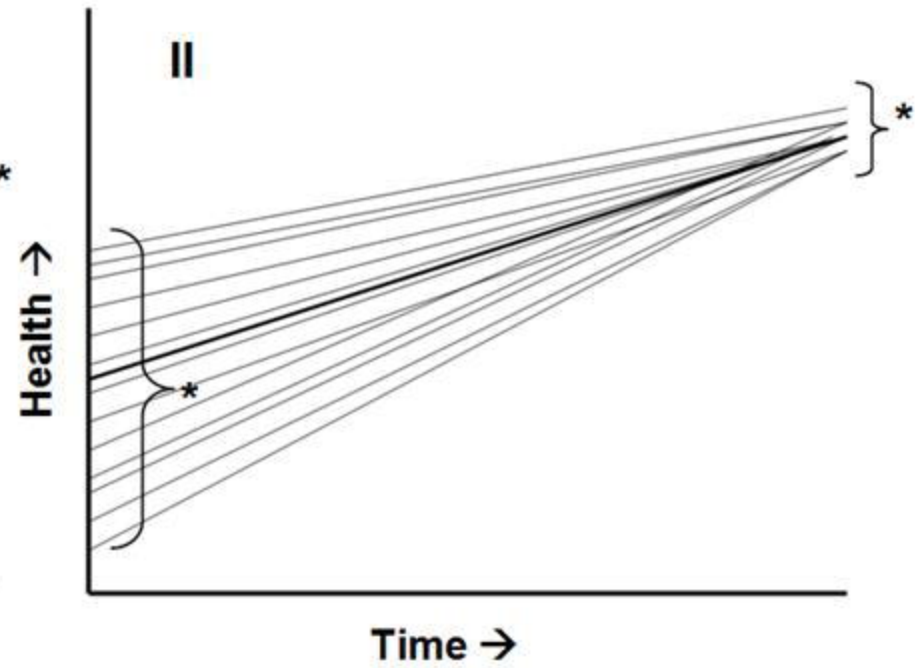
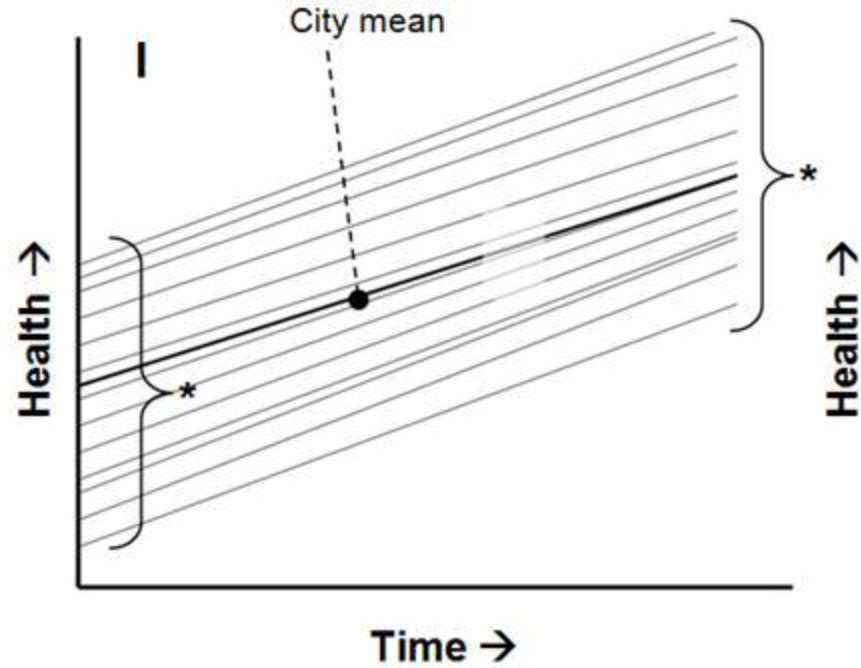
Figure 1. Visualization of a hypothetical association between a contextual variable (e.g., extreme poverty to extreme wealth) and individual **blood pressure**, based on data on from neighbourhoods in different cities around the world. In this representation the contextual variable explains 100% of the variance between neighbourhoods, since all areas are located on a regression line. However, while the regression coefficient is identical in all cities (i.e., a one unit increase in wealth results in an equivalent increase in health), the original neighbourhood variance before the introduction of the contextual variable was lower in city B than in city A. The original neighbourhood variance also corresponds to a much larger proportion of total individual variance in city A as opposed to city B. Consequently, a contextual public health intervention would be more efficient in city A than in city B. The key for this intervention is the size of the neighbourhood variance.

Figure 2. Visualization of an imaginary association between the interval of time after the start of a contextual intervention in a hypothetical city and mean individual health measured across (1) the entire city (thick line) and (2) in each of the city's neighbourhoods (thin lines). The measure of association (regression coefficient β) is positive; it is similar in I and II and shows an increase of the overall mean from the start of the intervention. However, health inequalities (i.e., neighbourhood variance) do not change along the time axis in scenario I, but are considerably reduced in scenario II.

The regression coefficient (β) is zero in both III and IV, which might be interpreted as an absence of contextual effects, since the intervention does not seem to influence health. However, while the health inequalities represented by neighbourhood variance have not changed in scenario III, they have been greatly reduced in scenario IV.

Individual health →





*Neighbourhood variance