LUND UNIVERSITY

Revisiting causal neighborhood effects on individual ischemic heart disease risk: A quasi-experimental multilevel analysis among Swedish siblings.

Merlo, Juan; Ohlsson, Henrik; Chaix, Basile; Lichtenstein, Paul; Kawachi, Ichiro; Subramanian, S V

Link to publication

Total number of authors:
6

Manuscript title

Revisiting causal neighborhood effects on individual ischemic heart disease risk: a quasi-experimental multilevel analysis among Swedish siblings

Author list and affiliations in the correct order

Juan Merlo
Henrik Ohlsson
Basile Chaix
Paul Lichtenstein
Ichiro Kawachi
SV Subramanian

Juan.merlo@med.lu.se
Henrik.Ohlsson@med.lu.se
chaix@u707.jussieu.fr
Paul.Lichtenstein@ki.se
IKAWACHI@hsph.harvard.edu
SVSUBRAM@hsph.harvard.edu

Social Epidemiology, Department of Clinical Sciences, Faculty of Medicine, Lund University, Sweden (Juan Merlo).

Department of Society, Human Development and Health, Harvard School of Public Health, Boston, USA (Ichiro Kawachi, Subramanian SV).

Unit for Primary Health Care Research, Region Skåne, Sweden (Juan Merlo, Henrik Ohlsson)

Centrum for Economic Demography, Lund University, Sweden (Juan Merlo)

Inserm, U707, Paris, France (Basile Chaix)

Université Pierre et Marie Curie-Paris6, UMR-S 707, Paris, France (Basile Chaix)

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (Paul Lichtenstein)

Corresponding author name, affiliation postal address and email address

Juan Merlo
Social Epidemiology, Faculty of Medicine, Lund University
Skåne University Hospital, Jan Waldenströms street 35
20502 Malmö
Sweden
Juan.merlo@med.lu.se

ABSTRACT

Neighborhood socioeconomic disadvantage is associated to increased individual risk of ischemic heart disease (IHD). However, the value of this association for causal inference is uncertain. Moreover, neighborhoods are often defined by available administrative boundaries without evaluating in which degree these boundaries embrace a relevant socio-geographical context that condition individual differences in IHD risk. Therefore, we performed an analysis of variance, and also compared the associations obtained by conventional multilevel analyses and by quasi-experimental family-based design that provides stronger evidence for causal inference

Linking the Swedish Multi-Generation Register to several other national registers, we analyzed 184,931 families embracing 415,540 full brothers 45 to 64 years old in 2004, and residing in 8,408 small-area market statistics (SAMS) considered as "neighborhoods" in our study. We investigated the association between low neighborhood income (categorized in groups by deciles) and IHD risk in the next four years. We distinguished between *family mean* and *intrafamilial-centered* low neighborhood income, which allowed us to investigate both unrelated individuals from different families and full brothers within families. We applied multilevel logistic regression techniques to obtain odds ratios (OR), variance partition coefficients (VPC) and 95% credible intervals (CI).

In unrelated individuals a decile unit increase of low neighborhood income increased individual IHD risk (OR = 1.04, 95% CI: 1.03−1.07). In the intrafamilial analysis this association was reduced (OR = 1.02, 95% CI: 1.02–1.04). Low neighborhood income seems associated with IHD risk in middle-aged men. However, despite the family-based design, we cannot exclude residual confounding by genetic and non-shared environmental factors. Besides, the low neighborhood level VPC= 1.5% suggest that the SAMS are a rather inappropriate construct of the socio-geographic context that conditions individual variance in IHD risk. In contrast the high family level VPC= 20.1% confirms the relevance of the family context for understanding IHD risk.

KEYWORDS

Sweden, Causality, Multilevel analysis, Myocardial ischemia, Residence characteristics, Siblings, Social medicine

The study of neighborhood effects on health is today an established field of research in social epidemiology (Diez Roux, 2001a; Kawachi & Berkman, 2003). In fact, numerous investigations try to identify causal associations between specific neighborhood social and economic characteristics and individual health conditions (Pickett & Pearl, 2001) such as ischemic heart disease (IHD) risk (Chaix B, 2009).

Investigating causality is a challenge to observational epidemiology, and the analysis of neighborhood causal effects on health presents specific difficulties (Oakes, 2004). Classic observational approaches try to reduce the effect of confounding by adjusting for numerous observable variables in multiple regression analyses, or by applying techniques like propensity scoring or inverse probability weighting (Hernan & Robins, 2006). However, the threat of residual confounding is difficult to eliminate. While randomized trials are the ideal study design for investigating neighborhood average causal effects (Oakes, 2004), they are also subject to many limitations, especially when the units of analysis are entire communities, rather than individuals (Diez Roux, 2004; Merlo & Chaix, 2006).

The methodological challenges inherent in the quantitative analysis of causal contextual effects are related to the identification of the relevant boundaries (Merlo et al., 2009), endogeneity, structural confounding, and extrapolation in multilevel regression analyses (Messer, 2007; Oakes, 2004). These difficulties have led some authors to conclude that knowledge of neighborhood effects could better be gained by qualitative approaches (Cummins et al., 2007).

However, new strategies of analysis, like quasi-experimental family-based designs (Lahey B.B. & D'Onofrio B. M., 2010), provide opportunities for investigating the effects of different contextual exposure in genetically related individuals (e.g., twins, full siblings) who also share a similar familial environmental background. This high intrafamilial correlation provides an opportunity to study the specific effects of a discrepant contextual exposure in

individuals who are—to some degree—interchangeable (e.g., full siblings residing in different neighborhoods with dissimilar degrees of deprivation). That is, the aim of a family-based design would be to approximate a counterfactual situation of exposure appropriate in analyzing causality. However, family-based designs do not completely eliminate issues of residual confounding because, besides remaining genetic differences, family members may also differ in life-course exposures to non-shared environmental factors that condition both the election of the place of residence and IHD risk. Nevertheless, the family design allows us to reduce the confounding effect of unobserved variables, which is a clear advantage compared to classic observational approaches.

Moreover the conventional study of *specific contextual effects* described above (i.e., the analysis of the association between specific neighborhood level variables and individual outcomes), we also intended to investigate *general contextual effects* that are estimated by measures of variance and clustering rather than by measures of association (Merlo, 2003; Merlo J et al., 2012; Merlo et al., 2009; Merlo et al., 2001; Petronis & Anthony, 2003; Subramanian SV et al., 2007). The study of general contextual effects shows the extent to which the geographical constructs we use for delimiting the neighborhood context (i.e., parishes, Swedish small-area market statistics, blocks, postal code areas, census tracts, etc.) may condition individual outcomes (e.g. IHD risk) without specifying any contextual characteristic other than the very boundaries we used for defining the context (Merlo et al., 2005). That is, if those boundaries really embrace a relevant socio-geographic context that conditions individual health, one should anticipate not only that there would be a statistically significant variation between neighborhood but also that this geographical variation represents a notable proportion of the total individual-level variation in the health outcome (Boyle MH & Willms JD, 1999; Duncan et al., 1993; Merlo, 2003; Merlo et al., 2004; Merlo et al., 2009; Subramanian SV et al., 2007) .

In summary, we sought to replicate previous observational findings suggesting that a discrepant exposure to neighborhood socioeconomic deprivation (e.g., low neighborhood income) is associated with individual IHD risk (Chaix B, 2009). For this purpose, we compared he results obtained by conventional multilevel analyses with those obtained when adopting a family-based design that considers full brothers nested within families. We analyzed "between" and "within" families associations. By design, the associations obtained by conventional multilevel analyses as well as those obtained in the between-families analyses are confounded for the same unobserved factors that are controlled for in the within-families associations. Therefore, while the between-families design resembles the classic multilevel analysis of neighborhood effects, the within-families estimations may be closer to a counterfactual situation of exposure, and therefore, provide more valid information for causal inference. Finally, our investigation explicitly aims to considering general contextual effects in the interpretation of specific contextual effects. In fact, without knowledge on general contextual effects, the specific contextual effects become "decontextualized" (Clarke P & Wheaton B, 2007; Merlo et al., 2009).

MATERIALS AND METHODS

Study population

In a first step, from the Swedish population residing in the country by December 31, 2004, we identified all the 971,519 men aged 45 to 64 years that fulfilled the inclusion criteria indicated below. We confined our investigation to men, as they have a higher incidence of IHD and more definite measures of socioeconomic position than women (Shavers, 2007). Moreover, we selected Swedish-born individuals, as the identification of familial connections among Swedish natives can be ascertained with far greater precision than those among immigrants

(Statistics Sweden, 2009). Thereafter, we excluded individuals previously hospitalized for IHD between January 1, 2001, and December 31, 2004. In addition, to improve the measurements of both individual and contextual exposures, we selected men residing in neighborhoods with at least 50 people and who had continuously resided in the country between December 31, 1992, and December 31, 2004. In a second step, by means of the Swedish Multi-Generation Register, we identified all full brothers and included families with two or more full brothers. This procedure yielded 184,931 families, including 43% (415,540/971,519) of the 45- to 64-year-old men.

Every resident of Sweden has a unique personal identification number that permits an accurate linkage of records to be made between different databases. The Swedish authorities replaced this number with an arbitrary code in order to protect the anonymity of the individuals in the study. Using this code, we linked several registers administered by the National Board of Health and Welfare (i.e., the National Patient Register and the National Cause of Death Register) and by Statistics Sweden (i.e., the Swedish Multi-Generational Register, the Register of Total Population, and the Longitudinal Integration Database for Health Insurance and Labor Market Studies). From these registers we obtained information on individual demographic and socioeconomic characteristics, hospitalizations, causes of death, and family and neighborhood variables.

The study was approved by the Ethics Committee at the Karolinska Institute (Dnr 2009-939).

Assessment of individual-level variables

The outcome variable in our study was IHD incidence, defined according to the International Classification of Diseases, 10th Revision, codes I20–I25. Between January 1, 2005, and December 31, 2008, we identified all individuals with the above discharge diagnosis in the Swedish Patient Register, or as underlying or contributing cause of death according to the

Causes of Death Register.

We dichotomized civil status in the year 2004 into living alone (i.e., single, divorced, or widowed) or not living alone (i.e., married, living in a registered partnership, or cohabiting and having children in common). We considered those not living alone as the reference group.

We defined unemployment history as being registered as unemployed in the years 1992, 1998, or 2004. We used the group that never experienced unemployment in any of these three years as a reference.

For each of the years 1992, 1998, and 2004 we computed decile groups of individualized disposable household income. Consequently, every individual obtained the value of his decile income group (i.e., 1, 2, 3… 10). Thereafter, we totaled the yearly values to obtain a variable ranging from 3 to 30. Subsequently, we reordered the cumulative income into groups by deciles (hereafter referred to as "individual income"). We performed this procedure for the whole Swedish population aged between 25 and 64. Therefore, this variable expresses the distribution in the country rather than in the study sample.

Assessment of neighborhood-level variables

We defined neighborhoods on the basis of small-area market statistics (SAMS) from Statistics Sweden. SAMS refers to the smallest administrative area units in Sweden. Those units have an average population of about 2,000 in Stockholm and about 1,000 in the rest of Sweden. The SAMS boundaries are drawn to include similar types of housing in a neighborhood. We recorded an individual's SAMS for every year from 1992 to 2004. For our analysis we identify 8,408 neighborhoods having at least 50 people.

We operationalized neighborhood socioeconomic circumstances by the proportion of low-income individuals between the ages of 25 and 64 years. For this calculation we used the

whole Swedish population. Therefore, this variable expresses the distribution in the country rather than in the study sample. For this purpose, we defined low individual income as the lowest quartile in the yearly individual income distribution. We calculated this proportion every year between 1992 and 2004 for each SAMS area. Thereafter, we calculated the deciles of the yearly distribution, so that each SAMS area obtained the value of its decile group (i.e., between 1 and 10) every year. Next, we totaled the yearly values for every individual and obtained a cumulative value for the period from 1992 to 2004 (i.e., between 13 and 130). Finally, we categorized the variable into deciles (hereafter referred to as "low neighborhood income"). This cumulative low neighborhood income variable accounts for the fact that the individuals may move from one place to another. In the appendix we provide a visual description of the construction of the cumulative low neighborhood income variable.

Using the values of the cumulative low neighborhood income for each individual, we also calculated a new neighborhood-level variable by averaging the individual values in each neighborhood. We denominate this variable *mean low neighborhood income*.

Assessment of family-level variables

We calculated *family mean low neighborhood income* as the average of the values of the variable representing low neighborhood income of the full brothers.

Subsequently, we computed *intrafamilial-centered low neighborhood income* as the difference between the value of the variable representing low neighborhood income of each sibling and the value of the variable representing familial mean low neighborhood income. A positive value indicates that this sibling is exposed to greater neighborhood deprivation than his full brothers.

All the low neighborhood income variables are comparable, as they have "decile groups" as

their measurement units. The association between decile groups of income and individual IHD risk was approximately linear and dose–response, and therefore, we modeled income as continuous variable in the analyses.

Statistical methods

We applied multilevel logistic regression analyses. This technique accounts for the hierarchical structure of the information with level 1 units (i.e., brothers) nested within level 2 units (e.g., families/neighbourhoods), and it estimates regression coefficients with a level 2-specific interpretation (Carlin et al., 2005). The multilevel logistic regression analysis allows calculating both *specific contextual effects* (i.e. the association between a particular area characteristic and individual risk) and *general contextual effects* (i.e. the degree to which the context, as a whole, conitions individual variance in mortality risk)

Specific contextual effects

We investigated the association between low neighborhood income and individual IHD risk. Firstly, we performed a conventional multilevel analysis with individual nested within neighborhoods in 2004. For this analysis we calculated the association between the mean neighborhood low income and individual IHD risk.

Secondly, we aimed to compare the results from the classic multilevel analysis of neighborhoods with the results from the sibling design. Therefore, we performed a sibling design and distinguished between familial mean low neighborhood income and intrafamilial-centered low neighborhood income. The familial mean low neighborhood income allows measuring of differences between families. Theoretically, this approach yields a confounded association since, by design, the exposure is calculated for groups (i.e., families) that are unrelated and differ from each other with regard to unknown genetic and shared

environmental factors. The intrafamilial-centered low neighborhood income was intended to approximate the counterfactual situation of exposure and allows contrasting of the IHD experiences of siblings living in different neighborhoods with dissimilar levels of low neighborhood income (Lahey B.B. & D'Onofrio B. M., 2010). The multilevel regression provides a family-specific regression coefficient for this variable that is adjusted for the familial cluster (i.e., the "family cluster" is included as a random term in the regression equation), and therefore, it accounts for an array of unknown shared genetic and environmental factors.

In the analyses we first estimated an age-adjusted model (model 1) followed by a model that also included the individual level variables living alone, unemployment, and income (model 2).

General contextual effects

General contextual effects are assessed by measures of variance and clustering rather than by measures of association (Merlo, 2003; Merlo J et al., 2012; Merlo et al., 2009; Merlo et al., 2001; Petronis & Anthony, 2003; Subramanian SV et al., 2007). We aimed to quantify the extent to which the SAMS we use for delimiting the neighborhood context conditioned individual IHD risk without specifying any contextual characteristic other than the very boundaries we used for defining the SAMS' context (Merlo et al., 2005). That is, if those boundaries really embrace a relevant socio-geographic context that conditions individual IHD risk, one should expect not only a statistically significant variation between SAMS but also that this geographical variation represents a relevant proportion of the total individual-level variation in individual IHD risk.

The neighborhood variance obtained from a multilevel logistic regression analyses can be translated into measures of heterogeneity, like the median odds ratio (Larsen & Merlo, 2005), and clustering, like the variance partition coefficient (W. Browne et al., 2005; Goldstein et

al., 2002; Li J et al., 2008; Merlo et al., 2006b). These measures provide fundamental information for understanding the potential relevance of concrete definitions of neighborhoods we are using (e.g., SAMS definition) as determinants of individual health disparities (Merlo et al., 2009). This information is, in turn, pertinent to public health policy when it comes to identifying the appropriate level of intervention (Merlo, 2003; Merlo et al., 2009).

We calculated the intra-class correlation using the latent variable method (W. Browne et al., 2005; Goldstein et al., 2002; Li J et al., 2008; Merlo et al., 2006b). This approach assumes that the propensity of the outcome is a continuous latent variable underlying our binary response. Each individual has a propensity for the outcome, but only people whose propensity exceeds a certain threshold will express the outcome. Using the logit link, it is assumed that the errors follow the standard logistic distribution and that the unobserved individual underlying variable follows a logistic distribution with individual variance equal to 3.29 ($\pi^2/3$).

$$\text{ICC} = \frac{\sigma_j^2}{\sigma_j^2 + \frac{\pi^2}{3}} \times 100$$

where σ_j^2 is the area variance, and $\frac{\pi^2}{3}$ is the individual-level variance.

For illustrative purposes we calculated the predicted probability of suffering an IHD event at the different decile groups of the neighborhood low-income variables. Using this information, we obtained the relative risk of suffering an IHD event, having the lowest decile group as reference (Figure 1).

The multilevel regression models were estimated with restricted iterative generalized least squares (RIGLS) followed by Markov chain Monte Carlo (MCMC) methods (W. J. Browne,

2009a) with parameter expansion at the family level to improve the estimation of the variance (W. J. Browne, 2009b). We compared models using the Bayesian deviance information criterion (DIC), and considered a reduction of the DIC greater than 10 as an indication of a better fit, adjusted for model complexity (Spiegelhalter et al., 2002). We extracted the median, 2.5%, and 97.5% values of the posterior distribution to calculate the point estimates of the parameters and their 95% credible interval (CI). We performed the analyses using MLwiN, A software package for fitting multilevel models, Centre for Multilevel Modeling, University of Bristol, version 2.23.

RESULTS

Characteristics of the population

Table 1 indicates the characteristics of the 45- to 64-year-old full brothers included in the study, as well as the characteristics of the population of men before the selection of the family groups. The information is stratified by low neighborhood income and age.

(Table 1 here)

The characteristics of the subpopulation of full brothers proved to be similar to men of the general population. It is clear that, overall, the percentage of men with low income, living alone, and being unemployed increased with every increment of cumulative low neighborhood income. The same pattern can be observed concerning IHD risk.

Specific contextual effects

Table 2 shows the results of the classic multilevel regression analysis with individuals nested within neighborhoods. A decile-unit increase in mean low neighborhood income increased

individual IHD risk (i.e., OR = 1.05, 95% CI: 1.04–1.05). However, this association became lower after adjustment by individual income, unemployment, and civil status in model 2 (OR = 1.03, 95% CI (1.03–1.04)).

Table 3, model 1, indicates that in unrelated individuals a decile unit of family mean low neighborhood income increased individual IHD risk (i.e., OR = 1.05, 95% CI: 1.04–1.06). However, when comparing full brothers using the intrafamilial-centered low neighborhood income, this association was weaker (OR = 1.04, 95% CI: 1.03–1.05), and it became further reduced after adjustment by individual income, unemployment, and civil status in model 2 (OR = 1.02 (1.02–1.04)).

Figure 1 represents the ratio between the predicted probability of suffering an IHD event at a specific decile group of low neighborhood income, and the predicted probability of suffering an IHD event at the first decile group (i.e., reference group). It shows that the strongest association between low neighborhood income and increased individual IHD risk appeared when we used the variable family mean low neighborhood income. On the other hand, we observed the weakest association when using the intrafamilial-centered low neighborhood income. The variable mean low neighborhood income (i.e., the classic multilevel analysis) produces an association of intermediate strength. In any case, adjustment for individual income, unemployment, and civil status further reduces the strength of the association between low neighborhood income and increased individual IHD risk.

General contextual effects

Tables 2 and 3 indicate that the family variance was about 16 times higher than the neighborhood variance. The value of the intra-class correlations was only 1.5% for the neighborhood level (Table 2), but as much as 20% for the family level (Table 3).

DISCUSSION

As we have discussed elsewhere (Merlo J, 2011; Merlo J et al., 2012; Merlo et al., 2009; Subramanian SV et al., 2007), from both an epidemiological and a public health policy viewpoint we need to consider two aspects in the investigation of neighborhood effects on health: (i) the validity of the association between neighborhood characteristics and individual IHD risk (i.e., specific contextual effects) and (ii) the relevance of the neighborhood boundaries −like those defining the Swedish SAMS− for understanding individual health disparities (i.e., general contextual effects)

Specific contextual effects

We observed an association between low neighborhood income and individual IHD risk. This observation was in line with previous findings obtained by traditional multilevel analyses and investigating different socioeconomic characteristics of the neighborhood (Chaix B, 2009; Pickett & Pearl, 2001; Sundquist et al., 2004). However, the observed association decreased as we accounted for confounding factors using a quasi-experimental family-based design and adjustment for individual socioeconomic characteristics.

The observed association was strongest when using the family mean low neighborhood income that deliberately forced a confounded comparison of unrelated individuals from families that differ in unknown—and thereby difficult to control—shared genetic and environmental confounding factors. As expected, on the other hand, we detected a weaker association when applying the quasi-experimental family-based design using the intrafamilial-centered low neighborhood income (i.e., comparing full brothers with different contextual exposures). In fact, the intrafamilial design adjusts for unmeasured factors similar to those confounding the association between family mean low neighborhood income and IHD risk. Nevertheless, adjustment by known individual characteristics (i.e., income, unemployment,

and civil status) further reduced the magnitude of the association OR = 1.02 (1.02–1.04). Therefore, it seems that the sibling analysis alone did not allow us to completely disentangle neighborhood from individual level effects in this population of adult full brothers.

The quasi-experimental family-based design provides observational epidemiology with a powerful strategy for investigating causality (D'Onofrio B. M. et al., 2010; Goodnight J.A. et al., 2010; Lahey B.B. & D'Onofrio B. M., 2010 ; Merlo, 2010). This design may also be useful in the analyses of neighborhood effects on health (Goodnight J.A. et al., 2010). However, we need to consider that in addition to discrepancies in the contextual exposure of interest, full brothers still have a 50% difference in genetic background. Full brothers also differ in non-shared environmental circumstances acquired across their life course, and these circumstances may be a common cause of both increased IHD risk and the selection of a neighborhood of residence.

In our study we considered the duration of each siblings exposure to low neighborhood income levels over the last 13 years. However, the individuals studied were between 32 and 51 years old when we began the investigation, and by that age siblings may have already moved to areas as conditioned by individual characteristics acquired as adults. To control for this aspect, we also adjusted for individual socioeconomic position (i.e., income and unemployment) and civil status, and we restricted the study sample to men without previous IHD hospitalizations. Even so, this strategy might have been insufficient, and thus we cannot exclude the possibility that the association observed in adults reflects the effect of residual compositional confounding.

A further limitation of our study is the lack of clinical data on individual risk factors. In any case, it can be discussed whether clinical conditions can be conceptualized as confounding or mediating factors (or both) of the effect of individual and of neighborhood socioeconomic

characteristics on IHD risk (Diez Roux, 2001b). Moreover, neighborhoods (i.e., SAMS areas) with similar socioeconomic circumstances may differ in other environmental characteristics that influence individual IHD risk. This situation could also be an additional source of residual confounding.

We performed a logistic regression rather than a survival analysis. However, the results were very similar using a family-stratified conditional Cox regression. This procedure, however, cannot estimate the association between familial mean low neighborhood income and individual IHD, since there is no variation within families in mean low neighborhood income.

Our sibling analysis, by including families with at least two full brothers, reduced the sample size and may have introduced selection bias. Nevertheless, the characteristics of the sibling subpopulation were very similar to those of the overall population.

The estimation of neighborhood causal effects is a cumbersome enterprise (Oakes, 2004) that convey specific difficulties absent when investigating causal effects at the individual level. A fundamental difference from individual-level analyses is that the geographical and administrative limits that define the neighborhoods (e.g., SAMS areas) do not necessarily represent the contexts that affect the individuals (Cummins et al., 2007; Merlo et al., 2009). This aspect needs be investigated by analyzing general contextual effects

General contextual effects

The structure of the information in our study was cross-classified with individuals grouped within both families and multiple neighborhoods over time (Browne WJ et al., 2001). However, the clustering of individual IHD risk within neighborhoods in 2004 was very low (i.e., ICC = 1.5%), and we have recently shown similar results in Sweden using parishes as a proxy of neighborhoods and taking into account the residential history of individuals in a cross-classified multilevel analysis (Ohlsson & Merlo, 2011). Consequently, the

neighborhood level can be disregarded without affecting the estimations of standard errors of the regression coefficients, and the low neighborhood income variable was analyzed as an individual-level variable. Obviously if the neighborhood level (i.e., the SAMS) can be disregarded in the multilevel regression analyses, this level should not be a central focus of investigation.

The low individual clustering of IHD risk within SMAS is not surprising, since the SAMS environment is only one small component among an array of spatial and cultural contexts to which individuals are exposed every day and throughout their lives. In fact, the SAMS appear as a very rough instrument for identifying individuals at risk. Therefore, knowledge on the size of the clustering of individual risk with neighborhoods is very pertinent in public health policy when it comes to identifying the appropriate level of intervention (Merlo, 2003; Merlo et al., 2009).

It is intuitive to assume that familial context is a major determinant of individual IHD risk, since full brothers not only have a similar heritability, but also share environments during critical periods in their lives (Lawlor DA et al., 2009; Merlo et al., 2006a). In fact, the intra-familial correlation in individual IHD risk was much higher (i.e., ICC = 20%) than the intra-neighbourhood correlation (i.e., ICC = 1.5%). It is just the existence of intra-familial correlation that qualifies the family design as quasi-experimental. Obviously, the higher the similarity between siblings (i.e., the higher the correlation within families), the better the family design is for investigating causality. In fact, if the siblings were fully interchangeable their similarity would be close to 100%. From this perspective, however, the correlation between full brothers was not so high (i.e., ICC = 20%). Therefore, the sibling design can be affected by confounding, as can any other observational analysis.

Conclusions

We compared full brothers who had different contextual exposures, and we were able to replicate the association between impaired socioeconomic characteristics in a neighborhood and increased individual IHD risk previously described (Chaix B, 2009). However, residual confounding by genetic and non-shared environmental factors may still be present.
In our study the neighborhood ICC was only 1.5%. Therefore, it seems that the SAMS are an inappropriate construct for capturing the relevant context that conditions individual IHD risk.

In summary, based on the unclear causal association between SAMS socioeconomic circumstances and individual IHD risk, together with the minor clustering of individual IHD risk within SAMS, we conclude that Swedish neighborhoods (as defined by the SAMS) play a minor role in conditioning individual IHD risk. However, it is known that Sweden is a highly egalitarian country with relatively low socioeconomic segregation (Islam et al., 2006), and therefore, our results may not be valid for other less egalitarian countries in the world- Possibly other neighborhood definitions might condition individual IHD risk in a higher degree than the SAMS do. In any case, from the perspective of the Swedish public health policy, rather than blaming some SAMS for their mean IHD risk, a national political system that improves the familial socioeconomic circumstances all over the country could be an effective strategy for preventing individual IHD risk.

REFERENCES

Boyle MH, & Willms JD (1999). Place effects for areas defined by administrative boundaries. *Am.J.Epidemiol.,* 149, 577-585.

Browne, W., Subramanian, S., Jones, K., & Goldstein, H. (2005). Variance Partitioning in Multilevel Logistic Models That Exhibit Overdispersion *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 168, 599-613.

Browne, W.J. (2009a). MCMC Estimation in MLwiN (Version 2.13) Centre for Multilevel Modelling, University of Bristol *http://www.bristol.ac.uk/cmm/software/mlwin/download/mcmc-print.pdf*.

Browne, W.J. (2009b). Parameter Expansion. . *MCMC Estimation in MLwiN (Version 2.13)* Bristol: Centre for Multilevel Modelling, University of Bristol

Browne WJ, Goldstein H, & Rasbash J (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling,* 1, 103-124.

Carlin, J.B., Gurrin, L.C., Sterne, J.A., Morley, R., & Dwyer, T. (2005). Regression models for twin studies: a critical review. *Int J Epidemiol,* 34, 1089-1099.

Chaix B (2009). Geographic Life Environments and Coronary Heart Disease: A Literature Review, Theoretical Contributions, Methodological Updates,and a Research Agenda. *Annu. Rev. Public Health,* doi:10.1146/annurev.publhealth.031308.100158.

Clarke P, & Wheaton B (2007). Addressing Data Sparseness in Contextual Population Research: Using Cluster Analysis to Create Synthetic Neighborhoods. *Sociological Methods Research,* 35, 311 - 351.

Cummins, S., Curtis, S., Diez-Roux, A.V., & Macintyre, S. (2007). Understanding and representing 'place' in health research: A relational approach. *Soc Sci Med,* 65, 1825-1838.

D'Onofrio B. M., Singh AL, Iliadou A, Lambe M, Hultman Ch M, Neiderhiser JM, et al. (2010). A quasi-experimental study of maternal smoking during pregnancy and offspring academic achievement. *Child Development (under 2nd revision)*.

Diez Roux, A.V. (2001a). Investigating neighborhood and area effects on health. *Am J Public Health,* 91, 1783-1789.

Diez Roux, A.V. (2001b). Investigating neighborhood and area effects on health. *American journal of public health,* 91, 1783-1789.

Diez Roux, A.V. (2004). Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Social Science & Medicine,* 58, 1953-1960.

Duncan, C., Jones, K., & Moon, G. (1993). Do places matter? A multi-level analysis of regional variations in health-related behaviour in Britain. *Soc Sci Med,* 37, 725-733.

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in generalised linear multilevel models. *Understanding Statistics,* 1, 223-232.

Goodnight J.A., Lahey B.B., Van Hulle C.A., Rodgers JL, Rathouz P.J., Waldman I.D., et al. (2010). A Quasi-Experimental Analysis of the Influence of Neighborhood Disadvantage on Child and Adolescent Antisocial Behavior. *Submitted to Journal of Abnormal Psychology on August 20, 2010*.

Hernan, M.A., & Robins, J.M. (2006). Estimating causal effects from epidemiological data. *J Epidemiol Community Health,* 60, 578-586.

Islam, M.K., Merlo, J., Kawachi, I., Lindstrom, M., & Gerdtham, U.G. (2006). Social capital and health: Does egalitarianism matter? A literature review. *Int J Equity Health,* 5, 3.

Kawachi, I., & Berkman, L.F. (2003). *Neighborhoods and health*. Oxford: Oxford University Press.

Lahey B.B., & D'Onofrio B. M. (2010). All in the Family: Comparing Siblings to Test Causal Hypotheses Regarding Environmental Influences on Behavior. *Current Directions in Psychological Science,* 19, 319-323.

Larsen, K., & Merlo, J. (2005). Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. *Am J Epidemiol,* 161, 81-88.

Lawlor DA, Mishra GD, & Eds (2009). *Family Matters. Designing, analysing and understanding family-based studies in life-course epidemiology*. New York: Oxford University Press.

Li J, Gray BR, & Bates DM (2008). An Empirical Study of Statistical Properties of Variance Partition Coefficients for Multi-Level Logistic Regression Models. *Communications in Statistics - Simulation and Computation,* 37, 2010–2026.

Merlo, J. (2003). Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *J Epidemiol Community Health,* 57, 550-552.

Merlo, J. (2010). (Book review) Family Matters: Designing, Analysing and Understanding Family-based Studies in Life-course Epidemiology. Debbie A Lawlor, Gita D Mishra (eds) *Int. J. Epidemiol. 39(3): 936-937 first published online January 16, 2010 doi:10.1093/ije/dyp387*

Merlo J (2011). Contextual Influences on the Individual Life Course: Building a Research Framework for Social Epidemiology. http://bit.ly/i5s64N. *Psychosocial Intervention,* 20, 109-118.

Merlo J, Viciana-Fernández FJ, & Ramiro-Fariñas D (2012). Bringing the individual back to small-area variation studies: a multilevel analysis of all-cause mortality in Andalusia, Spain. *Social Science & Medicine,* In press.

Merlo, J., Asplund, K., Lynch, J., Rastam, L., & Dobson, A. (2004). Population effects on individual systolic blood pressure: a multilevel analysis of the World Health Organization MONICA Project. *Am J Epidemiol,* 159, 1168-1179.

Merlo, J., Bengtsson-Bostrom, K., Lindblad, U., Rastam, L., & Melander, O. (2006a). Multilevel analysis of systolic blood pressure and ACE gene I/D polymorphism in 438 Swedish families--a public health perspective. *BMC Med Genet,* 7, 14.

Merlo, J., & Chaix, B. (2006). Neighbourhood effects and the real world beyond randomized community trials: a reply to Michael J Oakes. *Int J Epidemiol,* 35, 1361-1363.

Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., et al. (2006b). 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 1470-2738 (Electronic),* 60, 290-297.

Merlo, J., Chaix, B., Yang, M., Lynch, J., & Rastam, L. (2005). 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 1470-2738 (Electronic),* 59, 443-449.

Merlo, J., Ohlsson, H., Lynch, K.F., Chaix, B., & Subramanian, S.V. (2009). Individual and collective bodies: using measures of variance and association in contextual epidemiology. *J Epidemiol Community Health 1470-2738 (Electronic),* 63, 1043-1048.

Merlo, J., Ostergren, P.O., Hagberg, O., Lindstrom, M., Lindgren, A., Melander, A., et al. (2001). Diastolic blood pressure and area of residence: multilevel versus ecological analysis of social inequity. *J Epidemiol Community Health,* 55, 791-798.

Messer, L.C. (2007). Invited commentary: Beyond the metrics for measuring neighborhood effects. *Am J Epidemiol,* 165, 868-871; discussion 872-863.

Oakes, J.M. (2004). The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med,* 58, 1929-1952.

Ohlsson, H., & Merlo, J. (2011). Place effects for areas defined by administrative boundaries: a life course analysis of mortality and cause specific morbidity in Scania, Sweden. *Soc Sci Med,* 73, 1145-1151.

Petronis, K.R., & Anthony, J.C. (2003). A different kind of contextual effect: geographical clustering of cocaine incidence in the USA. *J Epidemiol Community Health,* 57, 893-900.

Pickett, K.E., & Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J.Epidemiol.Community Health,* 55, 111-122.

Shavers, V.L. (2007). Measurement of socioeconomic status in health disparities research. *J Natl Med Assoc,* 99, 1013-1023.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J R Stat Soc B,* 64, 583-639.

Statistics Sweden. (2009). Multi-Generation Register. In S. Sweden (Ed.), BE96 Bakgrundsfakta. Befolknings- och välfärdsstatistik. Stockholm.

Subramanian SV, Glymour MM, & Kawachi I. (2007). Identifying causal ecologic effects on health: a methodological assessment. . *In: Galea S, ed. Macrosocial determinants of population health. New York: Springer Media,.*

Sundquist, K., Malmstrom, M., & Johansson, S.E. (2004). Neighbourhood deprivation and incidence of coronary heart disease: a multilevel study of 2.6 million women and men in Sweden. *J Epidemiol Community Health,* 58, 71-77.

Table 1. Characteristics of the 45- to 64-year-old population of Swedish-born males residing in the country from 1993 to 2004, having no previous hospitalization for ischemic heart disease, and living in one of 8,408 neighborhoods with at least 50 people in 2004. Information is presented by decile groups of low neighborhood income for all men and for those with at least two full brothers.

	Number of people		Age (mean years)		Individual low income (%)		Unemployed (%)		Living alone (%)		Ischemic heart disease (2004−2008) (%)	
	All males	Full brothers	All males	Full brothers	All males	Full brothers	All males	Full brothers	All males	Full brothers	All males	Full brothers
Cumulative neighborhood low income (1992–2004)												
Decile group 1	104,455	40,939	55.0	55.0	3	3	12	11	28	27	2.7	2.5
2	103,947	41,931	54.5	54.6	4	4	16	16	38	37	3.0	2.9
3	97,369	39,899	54.4	54.4	5	5	18	18	41	41	3.2	3.1
4	100,706	41,820	54.3	54.3	6	6	20	20	43	42	3.3	3.2
5	98,928	42,310	54.3	54.3	7	7	21	21	42	41	3.5	3.3
6	91,640	39,315	54.2	54.3	8	8	23	23	44	42	3.6	3.5
7	10,3095	45,653	54.2	54.2	10	10	25	25	45	44	3.7	3.5
8	90,347	40,391	54.2	54.2	12	12	26	26	45	44	3.7	3.4
9	98,329	44,917	54.2	54.2	15	15	26	26	44	43	3.6	3.6
Decile group 10	82,703	38,310	54.3	54.2	20	21	26	26	46	45	3.8	3.7
Total	*971,519*	*41,5485*										

Because the number of people in every stratum of low cumulative neighborhood income was very large, the point estimations were precise, and so we do not provide 95% confidence intervals.

Table 2. Multilevel logistic regression with individuals nested within neighborhoods in 2004. The analysis investigates the association between cumulative low neighborhood income (1992–2004) and individual ischemic heart disease (2005–2008) in the 415,540 45- to 64-year-old Swedish-born males who were residing in the country from 1993 to 2004, with no previous hospitalization for ischemic heart disease, and who were living in one of 8,408 neighborhoods with at least 50 people by 2004. All models are age-adjusted. Values are odds ratios (95% credible intervals), unless otherwise indicated.

	Model 1	Model 2
Mean low neighborhood income* (One decile group increase)	1.05 (1.04–1.05)	1.03 (1.03–1.04)
Individual variables		
– Unemployment (yes vs. no)		1.20 (1.16–1.22)
– Living alone (yes vs. no)		1.18 (1.14–1.22)
– Income (one decile group decrease)		1.03 (1.01–1.04)
Neighborhood variance (95% confidence interval)	0.052 (0.019–0.065)	0.051 (0.021–0.067)
Intra-neighborhood correlation	1.6%	1.5%
Deviance information criterion	116,416	116,177

*Using the values of the cumulative low neighborhood income (1992–2004) for each individual, we calculated a neighborhood level variable by averaging the individual values in each neighborhood.

Table 3.   Multilevel logistic regression with individual nested families (full brothers) in 2004. The analysis investigates the association between cumulative low neighborhood income (1992–2004) and individual ischemic heart disease (2005–2008) in the 415,540 45- to 64-year-old Swedish-born males with at least one full brother who were residing in the country from 1993 to 2004, with no previous hospitalization for ischemic heart disease, and who were living in one of 8,408 neighborhoods with at least 50 people by 2004. All models are age-adjusted. Values are odds ratios (95% confidence intervals), unless otherwise indicated.

	Model 1	Model 2
Low neighborhood income (One decile group increase)		
– Familial mean*	1.05 (1.04–1.06)	1.04 (1.03–1.04)
– Intrafamilial-centered**	1.04 (1.03–1.05)	1.02 (1.02–1.04)
Individual variables		
– Unemployment (yes vs. no)		1.22 (1.17–1.27)
– Living alone (yes vs. no)		1.19 (1.15–1.23)
– Income (one decile group decrease)		1.03 (1.02–1.04)
Family variance (95% credible interval)	0.840 (0.728–0.947)	0.827 (0.723–0.931)
Intra-family correlation	20.3%	20.1%
Deviance information criterion	114,733	114,538

*Average of the values of the cumulative low neighborhood income (1992–2004) of the full brothers. **Difference between the value of the variable representing cumulative low neighborhood income of each sibling and the value of the variable representing familial mean low neighborhood income.
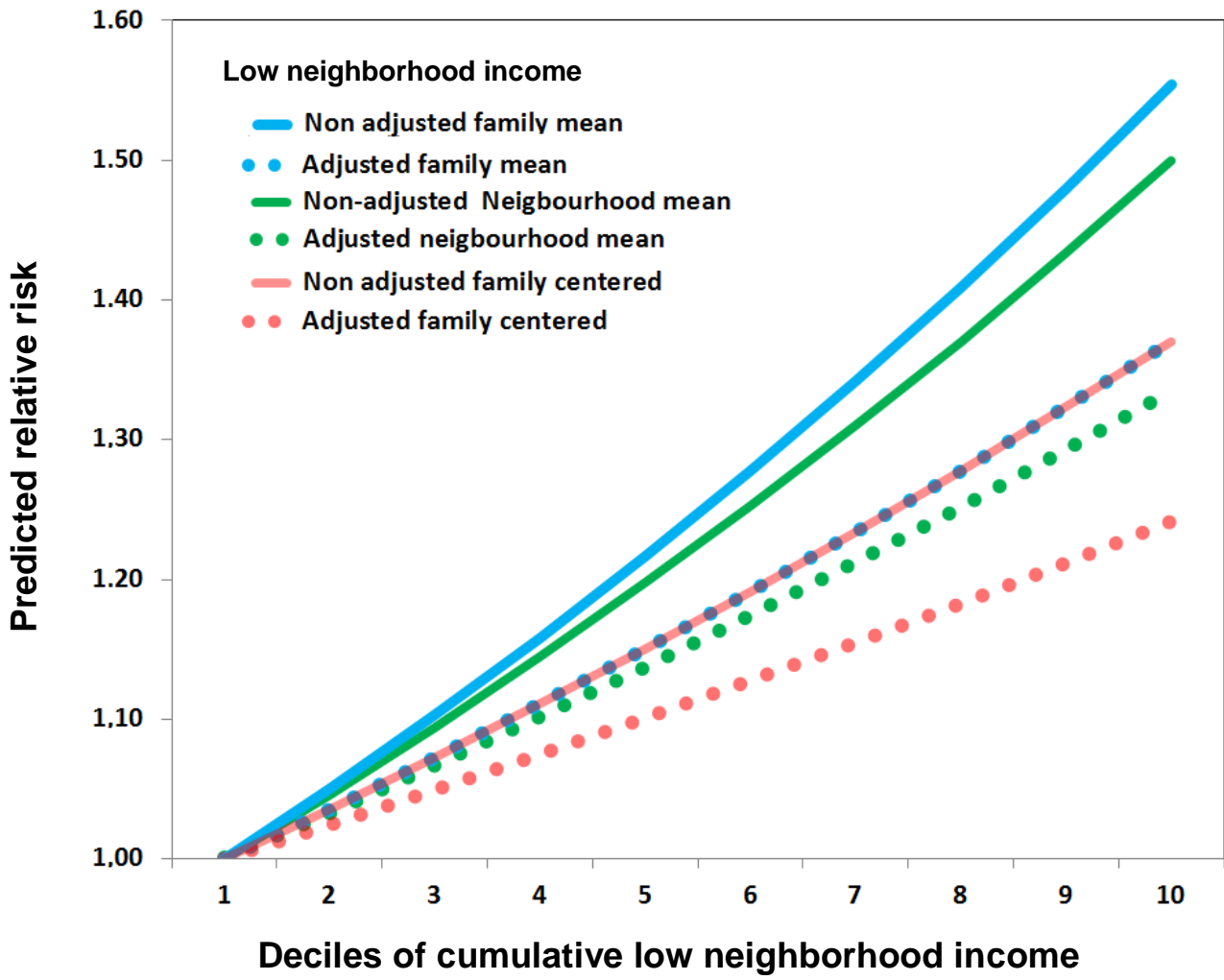
Figure 1

APPENDIX

Every year the neighborhoods are classed into decile groups according their percentage of people with low income (i.e., being within the first quartile of the individual distribution of income in the 25-65 year-old population)													Cumulative low neighborhood income	Deciles of cumulative low neighborhood income	
1	1	1	1	1	1	1	1	1	1	1	1	1	**13**	1	
2	2	2	2	2	2	2	2	2	2	2	2	2	**26**	2	
3	3	3	3	3	3	3	3	3	3	3	3	3	**39**	3	
4	4	4	4	4	4	4	4	4	4	4	4	4	**52**	4	**Sibling (full-brothers) design** • Family mean
5	5	5	5	5	5	5	5	5	5	5	5	5	**65**	5	• Family centered
6	6	6	6	6	6	6	6	6	6	6	6	6	**78**	6	
7	7	7	7	7	7	7	7	7	7	7	7	7	**91**	7	**Usual multilevel regression analysis**
8	8	8	8	8	8	8	8	8	8	8	8	8	**104**	8	• Neighborhood mean
9	9	9	9	9	9	9	9	9	9	9	9	9	**117**	9	
10	10	10	10	10	10	10	10	10	10	10	10	10	**130**	10	
Year 1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004			