Neighborhood-level Confounding in Epidemiologic Studies

Unavoidable Challenges, Uncertain Solutions

Basile Chaix,[a,b] *Cinira Leal,*[a,b] *and David Evans*[a,b,c]

Abstract: In early contextual studies, the aim was to demonstrate overall neighborhood influences rather than dissecting such influences into their components. Researchers did not need to worry about neighborhood-level confounding. However, as our interest shifts to the exploration of specific environmental effects, failure to consider neighborhood-level confounding may result in severely biased associations. We argue that neighborhood socioeconomic position and similar area structural factors may constitute powerful sources of confounding in studies of specific environmental factors and health. Controlling for neighborhood socioeconomic position is a convenient (but imperfect) adjustment strategy. Such control entails a minimal risk of overadjustment, but conveys a non-negligible risk of collider bias. Balancing the advantages and disadvantages, we suggest that researchers should often provide complementary analyses controlling for neighborhood socioeconomic position in studies of associations between specific environmental factors and health. Researchers should provide DAG-based descriptions of plausible scenarios to explore whether any decrease in the association of interest after adjustment for neighborhood socioeconomic position is likely due to neighborhood-level confounding, indirect pathway biases, or collider bias.

(*Epidemiology* 2010;21: 124–127)

Submitted 28 July 2008; accepted 4 September 2009; posted 10 November 2009.

From the [a]Inserm, U707, Research Unit in Epidemiology, Information Systems, and Modeling, Paris, France; [b]Université Pierre et Marie Curie, Paris6, UMR-S 707, Paris, France; and [c]EHESP School of Public Health, Rennes, France.

Supported by, as part of the RECORD project, the National Research Agency (Agence Nationale de la Recherche) (Health–Environment Program 2005 #00153 05), the Institute for Public Health Research (Institut de Recherche en Santé Publique), the National Institute for Prevention and Health Education (Institut National de Prévention et d'Education pour la Santé) (Prevention Program 2007 074/07-DAS), the National Institute of Public Health Surveillance (Institut de Veille Sanitaire) (Territory and Health Program), the French Ministries of Research and Health (Epidemiologic Cohorts Grant 2008), the National Health Insurance Office for Salaried Workers (Caisse Nationale d'Assurance Maladie des Travailleurs Salariés), the Ile-de-France Health and Social Affairs Regional Direction (Direction des Affaires Sanitaires et Sociales d'Île-de-France), the Ile-de-France Public Health Regional Group (Groupement Régional de Santé Publique), and the Ile-de-France Youth and Sports Regional Direction (Direction Régionale de la Jeunesse et des Sports).

Correspondence: Basile Chaix, INSERM U707, Faculté de Médecine Saint-Antoine, 27 rue Chaligny, 75012, Paris, France. E-mail: chaix@u707.jussieu.fr.

ISSN: 1044-3983/10/2101-0124
DOI: 10.1097/EDE.0b013e3181c04e70

In a recent series of observational studies,[1–4] Diez-Roux et al reported associations between specific characteristics of the neighborhood environment (eg, availability of healthy foods, walking environment, esthetic quality, safety, and social cohesion) and risk factors of coronary heart disease. These studies dealt with individual-level confounding but did not consider the possibility of neighborhood-level confounding.

In early contextual studies,[5,6] using neighborhood socioeconomic position (SEP) as the most common predictor, the aim was to demonstrate overall neighborhood influences rather than to dissect such influences into their components; thus, researchers did not need to worry about neighborhood-level confounding. However, as our interest shifts to the exploration of more specific environmental effects, failure to consider neighborhood-level confounding may result in severely biased associations.

We agree with Diez-Roux that a deeper understanding of environmental effects will require a combination of quantitative observational studies, qualitative approaches, interventional studies, and complex systems modeling.[7] Specifically, the present essay focuses on a methodologic issue relevant to quantitative observational research, assuming that refined analytic designs will help converge toward more informative associations.[8] The aim in this commentary is to emphasize methodologic concerns related to neighborhood-level confounding by highlighting how neighborhood SEP and other neighborhood structural factors may constitute powerful sources of confounding and by discussing the benefits and potential risks of adjustment for neighborhood SEP.

Neighborhood SEP as a Major Source of Neighborhood-level Confounding

SEP indicators such as income or wealth are "means of appropriation" of resources. Thus, a wide range of resources is accessible in high SEP neighborhoods, while harmful exposures are concentrated in low SEP neighborhoods. Figure 1 illustrates[9] a general pattern in which neighborhood SEP causally determines a number of exposures/resources, which can induce neighborhood-level confounding because of the resulting spatial covariations among exposures/resources. Thus, compared with more specific environmental factors, neighborhood SEP and similar area structural characteristics (such as degree of urbanicity) may be particularly important generators of neighborhood-level confounding.

FIGURE 1. Neighborhood SEP as an important source of neighborhood-level confounding when estimating the association between the environmental exposure of interest (in bold) and the outcome. A peculiarity of directed acyclic graphs (DAGs) in this case is that, even if all variables are intended to reflect individual-level constructs (including individual exposures to the environment), macro-macro processes, macro-micro processes, and micro-micro processes are potentially all involved. Also, these DAGs remain voluntarily imprecise on the exact temporal sequence of the processes, eg, whether the macro processes linking the various neighborhood exposures intervened before or after an individual moved into her/his neighborhood. (We verified that better specified DAGs incorporating these aspects led to comparable conclusions).

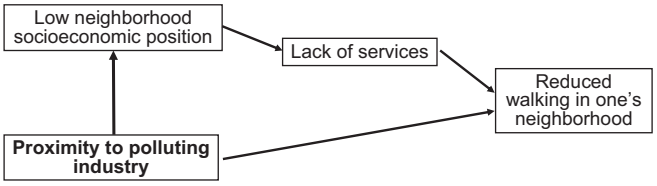

FIGURE 2. Neighborhood SEP on an indirect pathway biasing the association of interest between the environmental exposure (in bold) and the outcome.

However, it also implies that neighborhood SEP and similar factors can serve as convenient neighborhood adjustment factors that can allow researchers to control for a large number of resources/exposures (but of course the adjustment is imperfect when environmental exposures remain correlated conditional on neighborhood SEP).

Moreover, as there is often little reason to believe that neighborhood SEP intervenes in the causal pathway of a specific environmental effect, adjusting for neighborhood SEP entails only minor risks of overadjustment. This holds true in the situation in Figure 2: even if a polluting industry truly decreases neighborhood SEP over time through selective migration, the effect of the industry in which we are interested probably should not include the derived consequences of neighborhood SEP (eg, subsequent desertification of services or criminality). That is, environmental effects of interest should often be estimated independent of their ecologic side effects which, referring to the indirect effect pathway in Figure 2, may be considered as biasing pathways rather than pertinent mediating pathways. (Researchers of course need to pay attention to the assumptions for correctly estimating direct effects.[10]) Overall, adjustment for neighborhood SEP is therefore useful for mitigating both confounding and indirect pathway biases, and is, in most cases, unlikely to remove part of the causal effect of interest.

Of course, one should systematically check that the data allow adjustment of specific environmental effects for neighborhood SEP without excessive model extrapolations.[11] Also, future research designs should include a sufficient neighborhood-level sample size to deal with neighborhood-level confounding.

Potential Risks of Adjusting for Neighborhood SEP

A general pattern (Fig. 3) in which exposures/resources determine neighborhood SEP (eg, through selective migration) does not induce covariation among environmental variables and so does not induce neighborhood-level confounding. However, it does create a risk of collider bias upon adjustment for neighborhood SEP. (A collider is a variable with 2 arrows pointing into it.[9]) In Figure 3, while proximity to a polluting infrastructure and low-green space accessibility are marginally independent, the absence of one of these environmental determinants of neighborhood SEP in a low SEP neighborhood increases the probability that the other is present, creating a correlation between those environmental determinants within neighborhood SEP strata.[12] Thus, adjustment for neighborhood SEP would generate biases that were not initially present in the data. Of course, as in Figure 4, there are situations in which neighborhood SEP could induce an indirect biasing pathway if we do not condition on it, and a collider bias if we do.

FIGURE 3. Neighborhood SEP as a source of collider bias in the relationship between the environmental exposure of interest (in bold) and the outcome. (The collider is only shaped by the exposure).

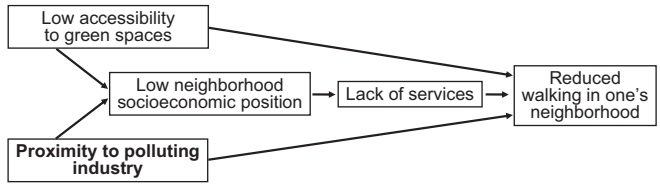

FIGURE 4. Neighborhood SEP on an indirect pathway biasing the association of interest between the environmental exposure (in bold) and the outcome, and as a source of collider bias when controlled for.

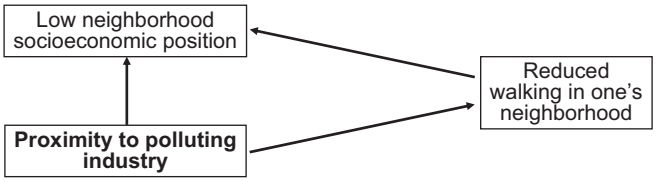

FIGURE 5. Neighborhood SEP as a source of collider bias in the relationship between the environmental exposure of interest (in bold) and the outcome. (The collider is causally shaped by both the exposure and the outcome).

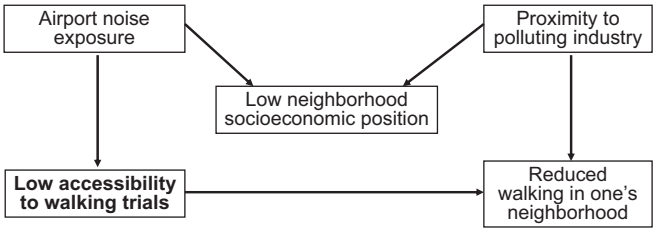

FIGURE 6. Neighborhood SEP as a source of collider bias in the relationship between the environmental exposure of interest (in bold) and the outcome. (The collider is causally shaped by neither the exposure nor the outcome).

A different collider bias is that shown in Figure 5, in which the outcome selectively encourages (or constrains) individuals to move to neighborhoods with a specific SEP (eg, life values causing the association between walking and neighborhood SEP). Adjustment for neighborhood SEP would create an association between polluting infrastructure and walking within neighborhood SEP strata, thus biasing the estimate of interest.

Figure 6 shows a more complex collider bias in which neighborhood SEP as the collider is affected by neither the main exposure (as in Figs. 3–5) nor the outcome (as in Fig. 5). In this so-called M-bias,[12] adjustment for neighborhood SEP could also affect the association of interest.

According to Greenland,[12] the bias is expected to be largest in Figure 5 (where both the exposure and the outcome affect neighborhood SEP), smaller in Figure 3 (where neighborhood SEP is shaped by the exposure but not by the outcome), and smaller still in Figure 6 (where neighborhood SEP is not affected by the exposure or outcome). However, the bias in Figure 5 requires important selective migration by individuals according to their outcome status, and is thus perhaps less likely than the bias in Figure 3.

Therefore, an important issue when evaluating whether adjustment for neighborhood SEP might generate collider bias is the direction of the association between the main environmental exposure and neighborhood SEP. Causal effects of specific environmental exposures/resources on neighborhood SEP are a likely source of neighborhood SEP-induced collider bias.

There is a situation—rather common in social epidemiology—in which collider bias from inappropriate neighborhood SEP adjustment would tend to move the estimate of an observed effect toward the null. Provided that the harmful environmental exposures that are the "parents" of neighborhood SEP are all positively correlated with low neighborhood SEP, conditioning on neighborhood SEP would pull any estimate of a true positive association between a harmful environmental exposure and an unfavorable health outcome toward the null or beyond, into the negative. In Figure 3, a neighborhood SEP-induced collider bias could exaggerate the environment–health association under study only if the environmental exposure of interest and the other environmental exposure (both parents of neighborhood SEP and negatively affecting health) were associated, one positively, the other negatively, with low neighborhood SEP. Conversely, in the common situation where environmental exposures that negatively affect the outcome all correlate positively with low neighborhood SEP, confounding could generate spurious associations, but collider biases would lead to either identifying an environmental effect under conservative conditions, masking an existing association, or producing a counter-intuitive association.

Of course, in cases where adjustment for neighborhood SEP tends to reduce the estimated environmental "effect" toward the null, deciding whether this change stems from removal of confounding or from introduction of collider bias will depend upon somewhat subjective judgments of the likelihoods of the various causal scenarios.

Dealing With Neighborhood-level Confounding: Unavoidable Challenges, Uncertain Solutions

As a general recommendation, attention should be more systematically given to neighborhood-level confounding in eco-epidemiologic studies, with DAG-based reasoning to conceptualize the risks involved.

One possible strategy is to adjust specific environmental effects for neighborhood SEP. In our view, adjustment for neighborhood SEP entails a minimal risk of overadjustment; it is often a convenient (but imperfect) adjustment strategy; it conveys a non-negligible risk of collider bias; but in the event of collider bias, such adjustment may also provide conservative estimates of the association of interest under certain (typically untestable) conditions.

Whether we should adjust a specific environmental effect for neighborhood SEP frequently appears as an epistemologic dilemma. We should keep in mind, however, that the extensively documented effect of area socioeconomic disadvantage on a large number of resources/exposures implies that neighborhood SEP may intervene as a major source of confounding in many epidemiologic studies.[1–4]

Epidemiologists would benefit from algebra- and simulation-based studies that quantify the likely patterns and

relative magnitude of confounding, indirect pathway biases, and collider biases arising from adjusting for neighborhood SEP. Future studies could explore the merits and drawbacks of various adjustment strategies for removing confounding while minimizing the risk of collider bias. Alternative approaches include direct adjustment for specific environmental confounders (rather than for generic variables such as neighborhood SEP), either included together in a single model, introduced in separate models, or aggregated into an environmental risk score.

In sum, complementary analyses controlling for neighborhood SEP should be provided, together with DAG-based descriptions of plausible scenarios examining whether any decrease in the association of interest upon adjustment for neighborhood SEP is likely due to neighborhood-level confounding, indirect pathway biases, or collider bias.

REFERENCES

1. Moore LV, Diez Roux AV, Nettleton JA, Jacobs DR Jr. Associations of the local food environment with diet quality–a comparison of assessments based on surveys and geographic information systems: the multiethnic study of atherosclerosis. *Am J Epidemiol*. 2008;167:917–924.
2. Mujahid MS, Diez Roux AV, Morenoff JD, et al. Neighborhood characteristics and hypertension. *Epidemiology*. 2008;19:590–598.
3. Mujahid MS, Diez Roux AV, Shen M, et al. Relation between neighborhood environments and obesity in the multi-ethnic study of atherosclerosis. *Am J Epidemiol*. 2008;167:1349–1357.
4. Auchincloss AH, Diez Roux AV, Brown DG, Erdmann CA, Bertoni AG. Neighborhood resources for physical activity and healthy foods and their association with insulin resistance. *Epidemiology*. 2008;19:146–157.
5. Riva M, Gauvin L, Barnett TA. Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations. *J Epidemiol Community Health*. 2007;61:853–861.
6. 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;30:81–105.
7. Diez Roux AV. Neighborhoods and health: where are we and were do we go from here? *Rev Epidemiol Sante Publique.* 2007;55:13–21.
8. 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.
9. Fleischer NL, Diez Roux AV. Using directed acyclic graphs to guide analyses of neighbourhood health effects: an introduction. *J Epidemiol Community Health.* 2008;62:842–846.
10. Kaufman JS, Maclehose RF, Kaufman S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiol Perspect Innov*. 2004;1:4.
11. Oakes JM. Commentary: advancing neighbourhood-effects research–selection, inferential support, and structural confounding. *Int J Epidemiol*. 2006;35:643–647.
12. Greenland S. Quantifying biases in causal models: classical confounding vs. collider-stratification bias. *Epidemiology*. 2003;14:300–306.