Neighborhood Effects on Health

Correcting Bias From Neighborhood Effects on Participation

Basile Chaix,[a,b] *Nathalie Billaudeau,*[a,b] *Frédérique Thomas,*[c] *Sabrina Havard,*[a,b] *David Evans,*[a,b,d] *Yan Kestens,*[e,f] *and Kathy Bean*[c]

Background: Studies of neighborhood effects on health that are based on cohort data are subject to bias induced by neighborhood-related selective study participation.

Methods: We used data from the RECORD Cohort Study (REsidential Environment and CORonary heart Disease) carried out in the Paris metropolitan area, France (n = 7233). We performed separate and joint modeling of neighborhood determinants of study participation and type-2 diabetes. We sought to identify selective participation related to neighborhood, and account for any biasing effect on the associations with diabetes.

Results: After controlling for individual characteristics, study participation was higher for people residing close to the health centers and in neighborhoods with high income, high property values, high proportion of the population looking for work, and low built surface and low building height (contextual effects adjusted for each other). After individual-level adjustment, the prevalence of diabetes was elevated in neighborhoods with the lowest levels of educational attainment (prevalence odds ratio = 1.56 [95% credible interval = 1.06–2.31]). Neighborhood effects on participation did not bias the association between neighborhood education and diabetes. However, residual geographic variations in participation weakly biased the neighborhood education–diabetes association. Bias correction through the joint modeling of neighborhood determinants of participation and diabetes resulted in an 18% decrease in the log prevalence odds ratio for low versus high neighborhood education.

Conclusions: Researchers should develop a comprehensive, theory-based model of neighborhood determinants of participation in their study, investigate resulting biases for the environment-health associations, and check that unexplained geographic variations in participation do not bias these environment–health relationships.

(*Epidemiology* 2011;22: 18–26)

Over the past 15 years, there has been a considerable development in the literature on neighborhood effects on health.[1–5] Cohort studies are typically used to investigate associations between neighborhood characteristics and health. However, such analyses suffer from a number of biases, including those related to selective participation in cohort studies,[6–8] which may distort the estimated associations between environmental exposures and health.[9]

As detailed in eAppendix 1 (section A1, http://links.lww.com/EDE/A434), many selective participation biases can be formulated in terms of collider bias.[10–12] When the environmental exposure and the outcome, or factors affecting the exposure or the outcome, have causal effects on study participation, participation intervenes as a collider (ie, a variable in a directed acyclic graph with at least 2 arrows pointing into it[11,13]). In these cases, conditioning on participation (in restricting the analysis to participants) can either generate an association between the environmental exposure and the outcome that does not exist in the source population or spuriously strengthen or weaken an existing association (eAppendix 1, section A1, http://links.lww.com/EDE/A434).[10,14] Because differential participation rates and loss of follow-up are observed even in epidemiologic cohorts recruited through random sampling,[7,15] researchers investigating neighborhood effects should systematically investigate neighborhood determinants of study participation.[10]
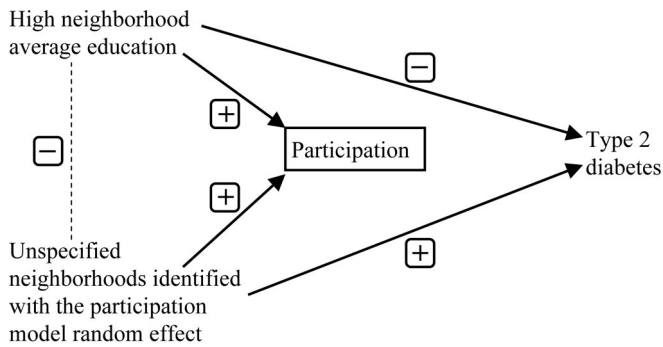
FIGURE 1. Unidentified neighborhood characteristics influencing participation in the study, if also associated with the outcome (type 2 diabetes), may bias the association of interest between neighborhood average education and diabetes. The dashed line represents the association generated by restricting the analyses to participants. Following Hernán et al,[10] the rectangle around participation indicates that the analyses condition on participation. The plus and minus signs indicate the direction of the associations observed in the data.

Our first aim was to develop a comprehensive, theory-based model of neighborhood determinants of participation in a cohort study on residential environment and coronary heart disease (the RECORD Cohort Study) (eAppendix 1, section A2, http://links.lww.com/EDE/A434). Our second aim was to examine whether neighborhood effects on study participation biased the associations between neighborhood socioeconomic variables and type 2 diabetes in this cohort (only a few previous studies have investigated relationships between neighborhood characteristics and diabetes[16]).

Biases in the environment–diabetes association may result either from the influence of identified neighborhood characteristics on study participation, or from the effects of unidentified neighborhood factors on participation (as illustrated in Fig. 1). We suggest that the neighborhood-level random effect of a model for study participation may be used to capture residual geographic variations in participation and control for its biasing effects. Building on Heckman selection models (eAppendix 1, section E6, http://links.lww.com/EDE/A434), we attempt to correct some of the selective participation biases through the joint modeling of neighborhood determinants of participation and neighborhood determinants of diabetes.

METHODS

Population

Our investigation of the neighborhood determinants of study participation relied on 2 distinct databases: (1) the RECORD Study database for the number of participants per neighborhood and their sociodemographic characteristics, and (2) the 1999 Census for the number of residents per neighborhood and their characteristics (denominators in the

analyses). Our study of the neighborhood correlates of diabetes was based on the RECORD Cohort.

The Cohort

A total of 7292 participants were recruited between March 2007 and February 2008. The participants were beneficiaries of the French National Health Insurance System for Salaried Workers, which offers a free medical examination every 5 years to all working and retired employees and their families (corresponding to 95% of the population of the Paris Ile-de-France region; eAppendix 1, section D1, http://links.lww.com/EDE/A434). Participants were recruited without a priori sampling during these 2-hour-long preventive checkups conducted by the Centre d'Investigations Préventives et Cliniques in 4 of its health centers, located in the Paris Ile-de-France region (Paris, Argenteuil, Trappes, and Mantes-la-Jolie). Eligibility criteria were as follows: age 30–79 years; ability to fill out study questionnaires; and residence in 1 of the 10 (out of 20) administrative divisions of Paris or 122 municipalities of the metropolitan area selected a priori (corresponding to a population of 5.2 million inhabitants in the 1999 Census). Among people presenting at the health centers who were eligible based on age and residence, 11% were not selected for participation because of linguistic or cognitive difficulties in filling out study questionnaires.[15] Of the persons selected for participation, 84% agreed to participate and completed the data collection protocol. Due to missing information, the available sample size was 7233 for study participation and 6876 for diabetes.

All participants underwent physical examination and filled out questionnaires. Participants were geocoded based on their residential address in 2007–2008. Research assistants rectified all incorrect or incomplete addresses with the participants by telephone. Extensive investigations with local Departments of Urbanism were conducted to complete the geocoding. Spatial coordinates and geographic codes of street, block, and block group were searched for each participant. Precise coordinates and block-group codes were identified for 100% of the participants. The study protocol was approved by the French Data Protection Authority.

The 1999 Population Census

The last available census (in 1999) was used for population denominators. A cross-tabulation provided the number of residents by age, sex, and education level for each neighborhood.

Individual and Neighborhood Measures
Analyses of Study Participation

The following individual characteristics were categorized the same way in both the Population Census and in the cohort study database: age (30–39; 40–59; and 60 years or older), sex, and education level (no education; secondary school and lower tertiary education; and higher tertiary education).

Neighborhoods were defined as census-block groups (IRIS areas in France). These were determined from the 1999 Census so as to be relatively homogeneous in sociodemographic and housing characteristics. Overall, 2218 neighborhoods were represented in the dataset matching the population Census to the cohort study database. Fewer neighborhoods were represented in the cohort study database (1882 neighborhoods for the analyses on diabetes), because there were no participants from several of the neighborhoods in the study territory. The median number of residents in the 2218 neighborhoods was 2264 in 1999 (interquartile range: 1959–2686). The median number of participants per neighborhood was 3 (interquartile range: 1–5). Neighborhood median area size was 0.16 km^2 (interquartile range: 0.08–0.35).

The following variables were considered at the neighborhood level: distance to the closest examination center; proportion of residents with a high education; median income; proportion of low-income residents not paying taxes; proportion of the active population looking for work; proportion of residents receiving social benefits; mean property value; population density; proportion of the area covered by buildings; mean building height; number of public transportation lines accessible in the neighborhood; density of services; ratio of specialty-care physicians to primary-care physicians; and an ecometric variable[17] for the degree of deterioration of the social/physical environment. Full details on these neighborhood variables and on hypotheses regarding their possible effects on study participation are reported in eAppendix 1, sections C1 and A3 (http://links.lww.com/EDE/A434). All environmental variables were divided into quartiles.

Analyses of Diabetes

Biologic parameters were measured under fasting conditions. Diabetes was defined as fasting blood glucose ≥126 mg/dL, or taking antidiabetic medication.

The following individual variables (described in eAppendix 1, section E2, http://links.lww.com/EDE/A434) were considered as possible correlates of diabetes: age and age squared, sex, marital status, education, and perceived financial strain. Three separate neighborhood variables (described in eAppendix 1, section C1, http://links.lww.com/EDE/A434) were used to characterize neighborhood socioeconomic position: the proportion of residents with a high education; median income; and mean property value (see eAppendix 1, section B, http://links.lww.com/EDE/A434 for hypotheses of neighborhood socioeconomic effects on diabetes).

Statistical Methods

Models for Study Participation

In the analyses of study participation, the outcome was the number of cohort study participants (ranging from 0 to 16) in each individual sociodemographic stratum (based on age, sex, and education) of each neighborhood from the preselected municipalities. We specified a Poisson-distrib-

uted error and a log link function. The logarithm of neighborhood population in the corresponding sociodemographic stratum in the 1999 Census was specified as the offset. Geographic variations in the rate of study participation were taken into account by including a neighborhood random effect in the model.

To assess spatial autocorrelation in study participation, we estimated the Moran's I statistic for the neighborhood random effect of the model. In the absence of spatial autocorrelation, the Moran's I statistic has a small negative expectation when applied to regression residuals.[18] To investigate whether spatial correlation decreased with increasing distance between locations, we computed Moran's I separately for neighborhoods less than 2000 meters apart, for those 2000–3999 meters apart, those 4000–5999 meters apart, and so forth.[19]

After estimating a model adjusted only for age and sex, we included individual education and the neighborhood variables in the model, retaining only those contextual variables that were independently associated with participation. We explored cross-level interactions between individual-level education and neighborhood variables. As recently recommended,[20] after testing a model incorporating a product term of ordinal variables for individual education and the neighborhood variable, we estimated a model with a 12-category variable combining categories of individual education and of the neighborhood variable (allowing us to examine whether there was an interaction on either the additive or the multiplicative scale).

As reported in eAppendix 1 (section F, http://links.lww.com/EDE/A434), we conducted a complementary analysis to distinguish between selection processes at different stages, ie, separate contextual influences on the rate of people going for a health checkup and contextual influences on study participation among subjects who went for the checkup.

Models for Diabetes

As detailed in eAppendix 1 (section E2, http://links.lww.com/EDE/A434), we developed a multilevel logistic model for diabetes, testing a number of individual and neighborhood sociodemographic explanatory variables. To identify potential participation-related collider biases, first we examined whether some of the neighborhood determinants of study participation were associated with diabetes. We then extracted the median of the posterior distribution of the random effect for each neighborhood from the model on study participation, and used this random effect divided into quartiles as an explanatory variable to assess whether residual geographic variations in study participation were associated with diabetes.

The random effect capturing residual geographic variations in participation is not a directly observed quantity, but rather a model estimate implying uncertainty. To

account for this uncertainty when estimating the association between residual geographic variations in study participation and diabetes, we used a Markov chain Monte Carlo approach to simultaneously estimate the model for the neighborhood determinants of study participation and the model for diabetes. In this joint modeling, at each iteration of the chain, the current values of the neighborhood random effect for study participation (different from one iteration to the next) are inserted as an explanatory variable in the model for diabetes, permitting the associations between neighborhood socioeconomic variables and diabetes to be adjusted more accurately for the somewhat-uncertain variable on rate of participation.

All models were estimated with Markov chain Monte Carlo simulation using WinBUGS 1.4.3.[21] All details on our estimation strategy are reported in eAppendix 1 (sections E3–E5, http://links.lww.com/EDE/A434) and the WinBUGS code for all models is reported in eAppendix 2 (http://links.lww.com/EDE/A434).

RESULTS

Models for Study Participation

A multilevel model adjusted for age and sex revealed important between-neighborhood variations in study participation. Based on the between-neighborhood variance (variance = 0.21 [95% credible interval = 0.18−0.25]), the rate of participation was 2.9 times higher (2.7–3.2) for the 25% of all residents in neighborhoods with the highest rates of participation compared with the 25% of all residents in neighborhoods with the lowest rates.[3,4,22] As shown with the Moran's I (Fig. 2), spatial autocorrelation in study participation was observed over a large range, but was modest in magnitude. The correlation decreased with increasing distance between neighborhoods, and vanished for neighborhoods 12 km or further apart.

The distribution of study participants and total population according to individual and neighborhood characteristics is reported in eAppendix 1 (section D2, http://links.lww.com/EDE/A434). A model containing individual and neighborhood variables indicated a markedly higher rate of study participation for those with high education attainment (Table 1). Rate of participation was lower for people residing far from the study center. Study participation was higher in both high median income and high mean property value neighborhoods after controlling for individual education. By contrast, participation was higher in neighborhoods with a high proportion of the active population looking for work.

Regarding physical environmental variables, independent associations indicated higher rates of study participation in neighborhoods with a low proportion of the area covered by buildings and a low mean building height. The ecometric variable representing the deterioration of the social/physical environment was not associated with participation. Pearson correlations between these neighborhood variables were moderate, with a few exceptions (eAppendix 1, section C2, http://links.lww.com/EDE/A434).

Product terms between individual education and neighborhood variables coded as ordinal variables indicated an interaction between the effects of individual education and distance to the center on the multiplicative scale. However, the model reported in Table 2 showed that the negative effect of distance on study participation was stronger among those with low education levels when assessed on the multiplicative scale; whereas the effect of distance was larger in the high-education group when the interaction was assessed on the additive scale.

As detailed in eAppendix 1, section F (http://links.lww.com/EDE/A434), complementary analyses conducted on people nested within municipalities confirmed that distance to the center and area indicators of

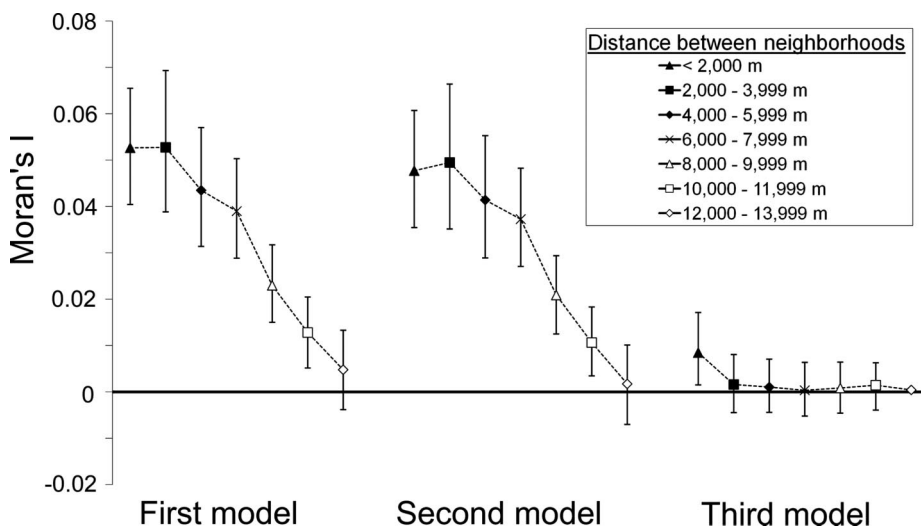FIGURE 2. Moran's I statistics and 95% credible intervals (vertical bars) for neighborhood-level residuals of multilevel models for participation in the cohort study, computed separately for pairs of neighborhoods less than 2000 meters apart, 2000−3999 meters apart, 4000−5999 meters apart, etc. The initial model included only age and sex; individual education and neighborhood factors were introduced in the second and third models.

TABLE 1. Associations Between Individual/Neighborhood Characteristics and Participation in the Cohort Study, as Estimated From a Multilevel Poisson Model (All Effects Adjusted for Each Other)

	Rate Ratio (95% Credible Interval)
Age (years)	
30–39[a]	1.00
40–59	1.84 (1.74–1.96)
60	1.37 (1.27–1.47)
Men (vs. women)	2.00 (1.90–2.10)
Individual education level	
Low[a]	1.00
Medium	1.90 (1.74–2.08)
High	4.25 (3.87–4.67)
Distance to the center	
High[a]	1.00
Mid-high	1.19 (1.09–1.30)
Mid-low	1.45 (1.32–1.58)
Low	1.75 (1.60–1.91)
Median income	
Low[a]	1.00
Mid-low	1.20 (1.09–1.32)
Mid-high	1.29 (1.14–1.45)
High	1.39 (1.20–1.60)
Mean property value	
Low[a]	1.00
Mid-low	1.10 (1.00–1.21)
Mid-high	1.11 (1.00–1.24)
High	1.23 (1.09–1.39)
Proportion of the active population looking for work	
Low[a]	1.00
Mid-low	1.01 (0.93–1.10)
Mid-high	1.18 (1.06–1.31)
High	1.31 (1.15–1.47)
Proportion of the area covered by buildings	
High[a]	1.00
Mid-high	1.13 (1.03–1.23)
Mid-low	1.26 (1.14–1.39)
Low	1.37 (1.23–1.51)
Mean building height	
High[a]	1.00
Mid-high	1.11 (1.03–1.21)
Mid-low	1.27 (1.16–1.39)
Low	1.27 (1.15–1.40)

[a]Reference category.

TABLE 2. Association Between Combined Categories of Individual Education and Distance to the Closest Center on the One Hand, and Participation in the Cohort Study on the Other Hand, Adjusted for Age, Sex, and Neighborhood Variables, as Estimated From a Multilevel Poisson Model[a]

Education Level and Distance to the Closest Center	Rate Ratio (95% Credible Interval)
Low education	
High distance[b]	1.00
Mid-high distance	1.23 (0.94–1.61)
Mid-low distance	1.56 (1.21–2.03)
Low distance	2.75 (2.19–3.47)
Intermediate education	
High distance	2.32 (1.93–2.83)
Mid-high distance	2.60 (2.15–3.19)
Mid-low distance	3.27 (2.71–4.02)
Low distance	4.06 (3.35–4.97)
High education	
High distance	5.28 (4.31–6.54)
Mid-high distance	6.61 (5.44–8.18)
Mid-low distance	7.49 (6.15–9.24)
Low distance	8.04 (6.59–9.91)

[a]On the multiplicative scale, the rate ratio for participation between people living nearby and far from the closest health center was 2.75 (2.75/1) in the low education group, 1.75 (4.06/2.32) in the intermediate education group, and 1.52 (8.04/5.28) in the high education group. In contrast, on the additive scale, for a base rate of participation equal to R, the effect of distance was 1.75R in the low education group, 1.74R in the intermediate education group, and 2.76R in the high education group.

[b]Reference category.

socioeconomic position and density were associated with going to the centers for health checkups, but were not associated (or associated only very weakly) with study participation among persons who were at the examination center for the health checkup.

In the final model for study participation, the between-neighborhood variance was reduced to 0.12 (95% credible interval = 0.09–0.14). As shown in Figure 2, spatial auto-correlation in study participation was to a large extent explained by the individual and neighborhood variables introduced into the model.

Models for Diabetes

As shown in Table 3 (first column), a low neighborhood education was associated with slightly higher odds of diabetes, after controlling for individual education and self-reported financial strain (see eAppendix 1, section E2 [http://links.lww.com/EDE/A434] for details on the construction of the model). Apart from neighborhood education, none of the neighborhood determinants of study participation (distance to the center, income, property value, proportion looking for work, building density, and height) showed associations with diabetes. Therefore, there was no need to adjust the model on diabetes for these neighborhood factors to remove participation-related collider biases.

The neighborhood-level random effect of the final model for study participation (capturing residual geographic variations in participation) was associated with the odds of diabetes, which were slightly higher in high-participation areas (Table 3, second column). The neighborhood random effect of the final model for participation showed almost no correlation with neighborhood education in the general population ($r = -0.004$ [95% confidence interval = -0.005 to

TABLE 3. Associations Between Individual and Neighborhood Characteristics and the Odds of Diabetes, as Estimated From Multilevel Logistic Models (All Effects Adjusted for Each Other), Before and After Controlling for Residual Geographic Variations in the Rate of Study Participation (n = 6876)

	Before Adjustment: Prevalence Odds Ratio (95% Credible Interval)	After Adjustment: Prevalence Odds Ratio (95% Credible Interval)
Age (1-year increase)	1.24 (1.07–1.38)	1.25 (1.13–1.41)
Age squared	1.00 (1.00–1.00)	1.00 (1.00–1.00)
Men vs. women	1.38 (1.05–1.84)	1.39 (1.06–1.86)
Living alone vs. cohabiting[a]	0.97 (0.72–1.30)	0.99 (0.73–1.32)
Individual education (vs. high)[a]		
Medium	1.40 (1.04–1.89)	1.39 (1.02–1.87)
Low	1.94 (1.26–2.92)	1.91 (1.24–2.88)
Perceived financial strain (vs. not)[a]	1.52 (1.07–2.14)	1.53 (1.07–2.16)
Neighborhood education (vs. high)[a]		
Mid-high	1.05 (0.70–1.56)	1.02 (0.68–1.52)
Mid-low	1.19 (0.80–1.75)	1.17 (0.79–1.73)
Low	1.56 (1.06–2.31)	1.50 (1.01–2.23)
Neighborhood random effect for study participation (vs. low)[a]		
Mid-low		1.19 (0.81–1.77)
Mid-high		1.31 (0.89–1.93)
High		1.58 (1.09–2.33)

[a]Reference category.

−0.002; n = 3.1 million). However, as expected from Figure 1, this random effect was negatively associated with neighborhood education in the sample of participants ($r = -0.14$ [95% confidence interval = −0.17 to −0.12; n = 7233). Compared with the general population, the relationship between the study participation random effect and neighborhood education was pulled into the negative in the sample of participants. Possibly this is because, if participation in the study is not caused by residing in a socially advantaged neighborhood, then it is likely that another cause of participation is present, eg, residing in one of these unspecified high-participation areas (identified from the participation random effect).

Due to this correlation, it is probably relevant to take into account residual geographic variations in study participation when estimating the association between neighborhood education and diabetes. As expected from Figure 1, the association between neighborhood education and diabetes was slightly reduced when the median of the posterior distribution of each neighborhood's participation random effect was introduced as a predictor in the model for diabetes (the change in effect size between the first and second columns of Table 3 was minimal but in the expected direction).

However, as noted above, the uncertainty associated with the random effect of the participation model would need

to be taken into account in our adjustment of the model for diabetes. To do so, we relied on the Markov chain Monte Carlo framework to estimate the model for the neighborhood determinants of study participation jointly with the model for diabetes (inserting the random effect of the first model as an explanatory variable in the second one) (Table 4). As shown in Table 4, in this joint model for participation and diabetes, the neighborhood random effect of the model for study participation was associated with the odds of diabetes. The log prevalence odds ratio for diabetes in low- versus high-education neighborhoods was 18% lower in the joint model (prevalence odds ratio = 1.44 [95% credible interval: 0.98–2.13]) than in the model of Table 3 (1.56 [1.06–2.31]), which does not control for residual geographic variations in study participation.

DISCUSSION

We found that a number of neighborhood factors related to the socioeconomic and physical environments were associated with participation in the RECORD Cohort Study, suggesting that participation biases may not depend only on individual characteristics but also on neighborhood features. Investigating associations between neighborhood socioeconomic variables and diabetes, we found that residual geographic variations in the rate of study participation were associated with diabetes. We attempted to correct the resulting bias in the relatively weak association between neighborhood education and diabetes that was observed through the joint modeling of the determinants of study participation and diabetes.

Strengths and Limitations

Strengths of the present study include a research design that allowed us to investigate individual/neighborhood determinants of participation in a cohort study, the large number of environmental correlates of participation that were available, the fact that residual random geographic variations in participation were conceptualized as a potential source of participation-related collider bias, and the joint-modeling framework implemented for bias correction.

One limitation of the participation analysis is the mismatch between the Census and the cohort study data. Discrepancies between numerators and denominators include the mismatch between the Census date (1999) and the cohort study recruitment dates (2007–2008), and the fact that individuals eligible for the health checkup had to be affiliated with the French National Health Insurance System for Salaried Workers, which corresponds to 95% of the total Census population. It is unlikely, however, that these small mismatches could have affected denominators of the participation rate enough to produce the observed associations with study participation. Another critical limitation is our inability to examine whether blood glucose or diabetes influenced study participation (we did not have information on diabetes

TABLE 4. Joint Modeling (i) of the Associations Between Individual and Neighborhood Characteristics and Participation in the Cohort Study, and (ii) of the Associations Between Individual and Neighborhood Characteristics and the Odds of Diabetes (All Effects in Each Model Adjusted for Each Other)

Participation in the RECORD Study	Rate Ratio (95% Credible Interval)
Age (vs. 30–39 years)[a]	
40–59 years	1.84 (1.73–1.95)
60 years and over	1.36 (1.27–1.47)
Men (vs. women)[a]	2.00 (1.90–2.10)
Individual education × distance to the center	
Low individual education	
High distance[a]	1.00
Mid-high distance	1.25 (0.96–1.63)
Mid-low distance	1.58 (1.22–2.04)
Low distance	2.77 (2.21–3.48)
Intermediate individual education	
High distance	2.36 (1.95–2.86)
Mid-high distance	2.63 (2.17–3.22)
Mid-low distance	3.30 (2.73–4.03)
Low distance	4.09 (3.38–5.00)
High individual education	
High distance	5.37 (4.38–6.61)
Mid-high distance	6.71 (5.49–8.25)
Mid-low distance	7.55 (6.20–9.27)
Low distance	8.10 (6.63–9.94)
Median income (vs. low)[a]	
Mid-low	1.19 (1.08–1.32)
Mid-high	1.28 (1.13–1.45)
High	1.39 (1.20–1.61)
Mean property value (vs. low)[a]	
Mid-low	1.10 (1.00–1.21)
Mid-high	1.12 (1.00–1.24)
High	1.24 (1.10–1.40)
Proportion of the active population looking for work (vs. low)[a]	
Mid-low	1.01 (0.93–1.10)
Mid-high	1.18 (1.07–1.32)
High	1.31 (1.16–1.49)
Proportion of the area covered by buildings (vs. high)[a]	
Mid-high	1.12 (1.03–1.21)
Mid-low	1.24 (1.13–1.37)
Low	1.34 (1.21–1.49)
Mean building height (vs. high)[a]	
Mid-high	1.11 (1.02–1.20)
Mid-low	1.25 (1.15–1.37)
Low	1.26 (1.15–1.38)

Diabetes	Prevalence Odds Ratio (95% Credible Interval)
Age (1-yr increase)	1.25 (1.12–1.37)
Age squared	1.00 (1.00–1.00)
Men (vs. women[a])	1.39 (1.05–1.85)
Living alone vs. cohabiting	0.98 (0.73–1.31)
Individual education (vs. high)[a]	
Medium	1.39 (1.03–1.88)
Low	1.88 (1.23–2.84)
Perceived financial strain (vs. not)[a]	1.52 (1.07–2.12)
Neighborhood education (vs. high)[a]	
Mid-high	1.01 (0.68–1.48)
Mid-low	1.15 (0.78–1.69)
Low	1.44 (0.98–2.13)
Neighborhood random effect for study participation (continuous)	2.90 (1.39–6.39)

[a]Reference category.

for either the general population or for persons who came to the health centers but did not participate in the study).

Main Findings

Neighborhood Influences on Study Participation

There were 3 selection stages in our recruitment strategy. First, populations attending the health centers are a selected sample. Second, particular participants were excluded by the staff (because of linguistic or cognitive limitations). Third, those who agreed to participate were also possibly nonrepresentative. Analyses reported in the main article amalgamated the 3 sources of selection, while those reported in eAppendix 1 (section F, http://links.lww.com/EDE/A434) distinguished between the first selection stage and an amalgamation of the second and third stages.

We found that the longer the road network distance to the closest health center, the lower the rate of study participation. As expected, complementary analyses confirmed that distance to the center predicted attendance for the health checkup, but that it did not predict study participation among people who were at the center for the health checkup (eAppendix 1, section F, http://links.lww.com/EDE/A434). Notably, the inhibiting effect of distance was more acute among persons with low education levels when assessed on the multiplicative scale, but weaker among those with low educational attainment when the interaction was assessed on the additive scale (due to the fact that the base rate of participation was much higher in educated than in noneducated participants). Because of the absence of strong theoretical guidance to decide whether the additive or multiplicative definition of effects should be used to gauge this interaction, it seems difficult to conclude firmly whether the distance effect on participation differed by education.

Independent of individual education effects, 2 mutually adjusted neighborhood effects (resulting from median income and

mean property value) indicated a lower rate of participation for residents of deprived neighborhoods.[15] Complementary analyses (eAppendix 1, section F, http://links.lww.com/EDE/A434) showed that low individual education did not strongly decrease the rate of people going for a health checkup, but that low education was strongly associated with low study participation among people who had come to the health centers for a checkup (perhaps reflecting a low interest in scientific studies[6,23] and exclusion because of linguistic or cognitive difficulties in filling out questionnaires among low socioeconomic groups). By contrast, a low neighborhood socioeconomic status was associated only with slightly lower rates of participation among people seen at the health centers, but was associated with a markedly lower rate of attendance at the centers (perhaps reflecting the spatial isolation of deprived neighborhoods, their lack of efficient public transportation, their residents' tendencies to rely on local resources, and collective norms that do not promote preventive healthcare). It is possible that we detected no neighborhood effect on participation among persons seen at the health centers because these people were a selected set who made significant efforts to attend the health centers. We cannot exclude that neighborhood effects would have been noted if a less selected population, contacted in its residential environment, were invited to participate.

By contrast, after adjustment, a high proportion of residents looking for work was associated with a higher rate of participation. As this variable reflects socioeconomic instability, a possible explanation is related to the specific recruitment targets of the participating health centers. Indeed, 3 of the 4 recruiting centers were set up in highly deprived areas specifically to reach patients with unstable economic resources.

Higher building density and building height were associated with lower rates of study participation.[15] In the absence of more convincing hypotheses, we can only speculate that residents of sparsely populated neighborhoods may have specific health-related attitudes encouraging them to attend preventive health examination centers.

Selective Study Participation as a Source of Bias in the Neighborhood–Diabetes Association

eAppendix 1 (section A, http://links.lww.com/EDE/A434) describes a number of situations in which environmental influences on study participation could bias environment–health associations. In our case, if building density (a determinant of participation) was a cause of diabetes, we would have to adjust for density in our analysis of neighborhood education and diabetes. This is because, even in the absence of a relationship between neighborhood education and building density in the general population, conditioning on participation would generate an association between them.

In our analyses, none of the identified neighborhood determinants of study participation could bias the relatively weak association between neighborhood education and diabetes. One of the original ideas of the study was to rely on the neighborhood random effect of the participation model to (1) capture the effects of unidentified neighborhood characteristics on study participation, (2) examine whether these residual geographic variations in participation were associated with diabetes, and (3) adjust the health model to remove a possible selective participation bias. By definition, this approach is not hypothesis-driven and does not need to be (we have no idea of the nature of neighborhood influences on participation captured by the random effect and why the latter was associated with diabetes). Overall, even if the bias correction leads only to a relatively weak change in the estimate of interest, our example illustrates that this strategy may enable a correction of participation-related collider biases that are not easily identifiable.

Implications for Future Investigations

Our study shows that it is feasible to investigate neighborhood determinants of participation in cohort studies. Of course, neighborhood-related selection may be much weaker when recruitment is based on a priori randomization and invitation of selected participants, and still weaker when participants are further surveyed and examined at home or nearby. However, relying on a randomized sample is not sufficient (due to selective nonparticipation and attrition), and epidemiologists, in addition to minimizing selection effects, should develop a comprehensive knowledge of the neighborhood determinants of participation in their study.

Overall, the general recommendations we make for ourselves, recommendations that may also be relevant for others, are as follows: (1) we will extend our analyses of the neighborhood determinants of participation in the RECORD Study; (2) we will rely on this comprehensive list of neighborhood determinants of participation to test their association with health outcomes in a search of participation-related collider biases; and (3) we will rely on the proposed joint modeling framework to verify that unexplained geographic variations in study participation do not bias the environment–health associations of interest.

REFERENCES

1. 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.
2. 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.
3. Chaix B, Rosvall M, Merlo J. Recent increase of neighborhood socioeconomic effects on ischemic heart disease mortality: a multilevel survival analysis of two large Swedish cohorts. *Am J Epidemiol*. 2007; 165:22–26.
4. Chaix B, Rosvall M, Merlo J. Neighborhood socioeconomic deprivation and residential instability: effects on incidence of ischemic heart disease and survival after myocardial infarction. *Epidemiology*. 2007;18:104–111.
5. Chaix B, Rosvall M, Merlo J. Assessment of the magnitude of geographic variations and socioeconomic contextual effects on ischaemic heart disease mortality: a multilevel survival analysis of a large Swedish cohort. *J Epidemiol Community Health*. 2007;61:349–355.
6. Goldberg M, Chastang JF, Leclerc A, Zins M, Bonenfant S, Bugel I. Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target population. *Am J Epidemiol*. 2001;154:373–384.
7. Nohr EA, Frydenberg M, Henriksen TB, Olsen J. Does low participation in cohort studies induce bias? *Epidemiology*. 2006;17:413–418.
8. Lissner L, Skoog I, Andersson K, et al. Participation bias in longitudinal studies: experience from the Population Study of Women in Gothenburg, Sweden. *Scand J Prim Health Care*. 2003;21:242–247.
9. Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol*. 1977;106:184–187.
10. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
11. 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.
12. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol*. 2010;39:417–420.
13. Chaix B, Leal C, Evans D. Neighborhood-level confounding in epidemiologic studies: unavoidable challenges, uncertain solutions. *Epidemiology*. 2010;21:124–127.
14. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.
15. Oakes JM, Forsyth A, Hearst MO, Schmitz KH. Recruiting participants for neighborhood effects research: strategies and outcomes of the Twin Cities Walking Study. *Environ Behav*. 2009;41:787–805.
16. Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obes Rev*. 2010 Mar 1. [Epub ahead of print.]
17. Chaix B, Lindstrom M, Merlo J, Rosvall M. Neighbourhood social interactions and risk of acute myocardial infarction. *J Epidemiol Community Health*. 2008;62:62–68.
18. Odland J. *Spatial Autocorrelation*. Newbury Park, CA:Sage Publications;1988.
19. Chaix B, Merlo J, Subramanian SV, Lynch J, Chauvin P. Comparison of a spatial perspective with the multilevel analytic approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. *Am J Epidemiol*. 2005;162:171–182.
20. Kaufman JS. Interaction reaction. *Epidemiology*. 2009;20:159–160.
21. Smith AF, Roberts GO. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J R Stat Soc Ser B Stat Methodol*. 1993;55:3–23.
22. Merlo J, Chaix B, Ohlsson H, 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.
23. Turrell G, Patterson C, Oldenburg B, Gould T, Roy MA. The socioeconomic patterning of survey participation and non-response error in a multilevel study of food purchasing behaviour: area- and individual-level characteristics. *Public Health Nutr*. 2003;6:181–189.