

# A Method to Detect Residual Confounding in Spatial and Other Observational Studies

W. Dana Flanders,<sup>a,b</sup> Mitchel Klein,<sup>c</sup> Lyndsey A. Darrow,<sup>c</sup> Matthew J. Strickland,<sup>c</sup> Stefanie E. Sarnat,<sup>c</sup> Jeremy A. Sarnat,<sup>c</sup> Lance A. Waller,<sup>b</sup> Andrea Winquist,<sup>c</sup> and Paige E. Tolbert<sup>c</sup>

**Background:** Residual confounding is challenging to detect. Recently, we described a method for detecting confounding and justified it primarily for time-series studies. The method depends on an indicator with 2 key characteristics: (1) it is conditionally independent (given measured exposures and covariates) of the outcome, in the absence of confounding, misspecification, and measurement errors; and (2) like the exposure, it is associated with confounders, possibly unmeasured. We proposed using future exposure levels as the indicator to detect residual confounding. This choice seems natural for time-series studies because future exposure cannot have caused the event, yet they could be spuriously related to it. A related question addressed here is whether an analogous indicator can be used to identify residual confounding in a study based on spatial, rather than temporal, contrasts.

**Methods:** Using directed acyclic graphs, we show that future air pollution levels may have the characteristics appropriate for an indicator of residual confounding in spatial studies of environmental exposures. We empirically evaluate performance for spatial studies using simulations.

**Results:** In simulations based on a spatial study of ambient air pollution levels and birth weight in Atlanta, and using ambient air pollution 1 year after conception as the indicator, we were able to detect residual confounding. The discriminatory ability approached 100% for some factors intentionally omitted from the model, but was very weak for others.

**Conclusion:** The simulations illustrate that an indicator based on future exposures can have excellent ability to detect residual con-

founding in spatial studies, although performance varied by situation.

(*Epidemiology* 2011;22: 823–826)

Residual confounding is difficult to detect, in part, because assessment must be based on causal considerations. It is not enough to model statistical associations accurately, as associations may not mirror causal patterns.<sup>1</sup>

We have previously proposed a method for detecting residual confounding or other forms of model misspecification based on the tenet that a cause must precede its effect.<sup>2,3</sup> The method uses a variable with 2 key characteristics. The variable should be independent of disease, absent confounding, or other misspecification, and it should be associated with the exposure and confounding covariates. Emphasizing time-series studies, we previously argued that future levels of many environmental exposures can often have the needed characteristics: future exposures cannot have caused past disease, and so an association with prior disease can be spurious.

Using future exposure variables in time-series studies to detect temporal confounding is intuitively attractive—future exposures cannot have caused prior health events and yet could be spuriously related to the event. A related question is whether an analogous variable can be used in spatial studies to identify residual confounding, such as in studies where region-specific pollutant concentrations are correlated with region-specific disease rates. Risk factors such as smoking could covary with pollutant concentrations across regions and thus lead to spatial confounding. Although the nature of confounding differs between spatial and time-series studies,<sup>4</sup> the previously proposed method for detecting residual confounding can be similarly applied in spatial studies. We present simulation results that evaluate the method's performance and discuss its strengths and weaknesses.

## METHODS

To illustrate the approach for spatial studies, we consider a study whose objective is to assess the association of area-specific air pollution levels (eg, ambient ozone,  $AP_0$  in the directed acyclic graph<sup>5,6</sup> in Fig. 1) with the birth weight of newborn infants in the same area ( $D_1$  in Fig. 1). An

Submitted 10 February 2011; accepted 17 May 2011.

From the Departments of <sup>a</sup>Epidemiology, <sup>b</sup>Biostatistics and Bioinformatics, and <sup>c</sup>Environmental and Occupational Health, Rollins School of Public Health, Emory University, Atlanta, GA.

Supported by EPA STAR RD83479901 and RD833626, NIEHS R01ES11294, and EPRI EP-P27723/C13172.

The views expressed in this document are solely those of the authors and do not necessarily reflect the views of the funding agencies, and mention of any products or commercial services does not constitute endorsement.

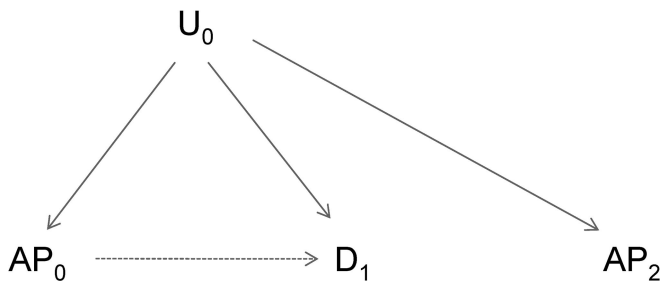
**SDC** Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article ([www.epidem.com](http://www.epidem.com)).

Correspondence: W. Dana Flanders, Department of Epidemiology, Rollins School of Public Health, Emory University, 1518 Clifton Rd NE, Atlanta, GA 30322. E-mail: [flanders@sph.emory.edu](mailto:flanders@sph.emory.edu).

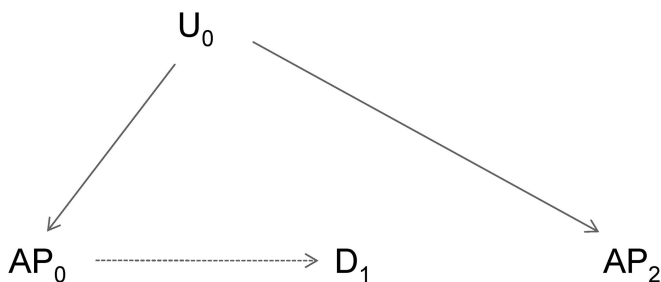
Copyright © 2011 by Lippincott Williams & Wilkins

ISSN: 1044-3983/11/2206-0823

DOI: 10.1097/EDE.0b013e3182305dac



**FIGURE 1.** Directed acyclic graph showing basic causal relationships among exposure (air pollution, AP<sub>0</sub>), the health outcome (birth weight, D<sub>1</sub>), an unmeasured covariate (U<sub>0</sub>), and future air pollution (AP<sub>2</sub>). Confounding by U<sub>0</sub> is present, indicated by the backdoor path from AP<sub>0</sub> to D<sub>1</sub>.



**FIGURE 2.** Directed acyclic graph showing basic causal relationships among exposure (air pollution, AP<sub>0</sub>), the health outcome (birth weight, D<sub>1</sub>), an unmeasured covariate (U<sub>0</sub>), and future air pollution (AP<sub>2</sub>). Confounding by U<sub>1</sub> is absent, as there is no open backdoor path from AP<sub>0</sub> to D<sub>1</sub>.

important presumption is that birth weight (D<sub>1</sub>) does not affect future pollution levels (AP<sub>2</sub>), reflected by the absence of an arrow from D<sub>1</sub> to AP<sub>2</sub>. Furthermore, spatial factors that affect health and are associated with future pollution levels (eg, poverty) should often also be associated with earlier pollutant levels, illustrated by U<sub>0</sub> (Fig. 1). If these relationships hold, we argue in the eAppendix 1 (<http://links.lww.com/EDE/A501>)—and previously for time-series studies<sup>3</sup>—that, conditional on AP<sub>0</sub>, future air pollutant levels should not be associated with disease in the absence of confounding (Fig. 2), but associated in its presence (Fig. 1). In this case, future pollutant levels have key characteristics needed for an indicator of unmeasured confounding<sup>2,3</sup> and can be used to detect it or other model misspecification.

The model used to analyze the study should include the exposure of interest (air pollutant level prior to outcome occurrence, AP<sub>a,i</sub>) and relevant covariates. Equation (1) illustrates a linear form:

$$E(Y_{a,i}) = \alpha + \beta * AP_{a,i} + \gamma * covariates_{a,i} \quad (1)$$

for infant *i* in area *a*, E(Y<sub>a,i</sub>) is the expected birth weight and AP<sub>a,i</sub> is a relevant, prenatal air pollutant level.

To use future area-specific air pollutant levels as an indicator, we also fit the same model that now also includes

the indicator variable (area-specific, pollutant level measured after infant *i*'s birth, say AP<sub>a,i</sub><sup>f</sup>):

$$E(Y_{a,i}) = \alpha + \beta * AP_{a,i} + \gamma * covariates_{a,i} + \delta * AP_{a,i}^f \quad (2)$$

If residual confounding and model misspecification are absent and the assumed causal relationships are adequate approximations, AP<sub>a,i</sub><sup>f</sup> should be unassociated with Y<sub>a,i</sub> after adjustment for covariates. An association between AP<sub>a,i</sub><sup>f</sup> and the outcome (δ ≠ 0) suggests residual confounding or other model misspecification. Thus, we use the statistic *I* to test for residual confounding:

$$I = \hat{\delta} / \hat{\sigma}_f \quad (3)$$

where  $\hat{\delta}$  is the estimated slope and  $\hat{\sigma}$  its standard error. Under the null hypothesis of no model misspecification, *I* is approximately normally distributed.

### Simulations

We assess the indicator's ability to detect residual confounding using data from a spatial study of ambient air pollution and birth weight in Atlanta. Use of simulations allows us to specify the “true” causal relationships; using estimated parameters to calculate the true, expected birth weights makes the simulations realistic. We consider 2 ambient pollutants commonly assessed in the birth outcomes literature, PM<sub>2.5</sub> (μg/m<sup>3</sup>) and NO<sub>2</sub> (parts per billion). We assess relationships between pollutant levels during the first month of pregnancy and birth weight of full-term infants. Zip code-specific levels for each pollutant were a weighted average estimate.<sup>7</sup>

Each full-term infant (37–43 weeks' gestation) was assigned to the Zip code of the mother's residence. We calculated the ambient air pollution level for each infant's Zip code, averaged over the 4 weeks after the estimated conception date. Modeled covariates included gestational age (weekly indicators); maternal education, age (linear spline with 3 knots), tobacco use (yes/no), Medicaid (yes/no), and race/ethnicity (non-Hispanic white, African-American, and Hispanic); and the Zip code's percent of population below the poverty line. We included indicators for date-of-conception in 2-week intervals, so comparisons were spatial.

We also calculated the air pollution level for each Zip code averaged over the 4-week period beginning 1 year after the conception date. Because these exposures occurred after birth (8+ weeks), they cannot have affected birth weight. Future levels (AP<sub>i,a</sub><sup>f</sup>) are included only in models that also include pollutant levels prior to birth and a subset of the covariates.

We used the model in Eq. (1), with Zip code as the area. We fit this model (once) to the observed birth weights to obtain model-predicted weights for each infant, and then treated the model-predicted weights as the true expected

values in the simulations. To assess the indicator's ability to detect confounding, we first generated a birth weight for each infant, using the true predicted values and adding a random Gaussian error. We fit the correct model including the indicator, and then, to simulate confounding, we fit a misspecified model (incorrectly omitting a factor) along with the indicator. In most scenarios, we omitted an actual, measured factor (eg, smoking). In a few situations, we created 3 hypothetical factors and included them when fitting the model to obtain alternative true predicted weights. We subsequently omitted one of the variables to simulate additional patterns of confounding. We calculated the proportion of simulations in which  $I$  exceeded 1.96 in absolute value, rejecting the null hypothesis of no confounding. We also calculated the area under the receiver operating curve as a measure of discriminatory ability. We include a program for simulating the power to detect model misspecification due to omission of a confounder (eAppendix 2, <http://links.lww.com/EDE/A504>). The user can either specify parameters to generate data hypothetically, or use actual observations to fit a model and base simulations on the fitted parameters.

## RESULTS

PM<sub>2.5</sub> was negatively associated with birth weight in these spatial analyses (Table 1, scenario 1;  $\hat{\beta} = 20.6\text{g}/10\ \mu\text{g}/\text{m}^3$ ). Compared with the true model that generated the data, improperly omitting various variables led to varying degrees of simulated confounding.  $\hat{\beta}$  changed by about 70% when age was omitted and by 700% when race was omitted (Table 1, column 3). The indicator's ability to detect simulated confounding also varied substantially. For the situations considered, the indicator's ability was weak when confounding was weak-to-modest, for example, when age or tobacco was omitted (AUC = 0.51–0.56, Table 1, column 5). However, this was sample-size dependent: with quadrupling of the sample size, the ability to detect confounding (created when the poverty variable is incorrectly omitted) increased from 19% (Table 1) to >50% (data not shown). With stronger degrees of simulated confounding, the indicator consistently signaled that confounding might be a problem (eg, scenarios 5 and 6; AUC = 0.93–1.00, Table 1).

Simulation results were similar for NO<sub>2</sub> (Table 2) and also when we considered confounding by the hypothetical

**TABLE 1.** Simulation Results

Scenarios	Median $\beta^a$ (SE ( $\hat{\beta}$ ))	Percent Bias	Median ( $I$ )	Proportion of Simulations in Which Null Is Rejected ( $H_0: \delta_f = 0$ )	AUC
1 (Correct model) <sup>b</sup>	-20.6 (17.2)	—	-0.01	0.050	0.50
2 (Drop tobacco)	0.0 (17.3)	100	0.64	0.086	0.56
3 (Drop age)	-5.8 (17.3)	72	0.19	0.042	0.51
4 (Drop percent below poverty line)	-50.6 (17.1)	146	-1.06	0.184	0.65
5 (Drop Medicaid, tobacco, age, education)	61.4 (17.4)	398	1.85	1.00	0.93
6 (Drop race)	-168 (17.3)	716	-5.48	1.00	1.00

True (correct) model stipulates an effect of exposure (PM<sub>2.5</sub>) on birth weight.

<sup>a</sup> $\beta$  is the estimated change in birth weight (g), per 10  $\mu\text{g}/\text{m}^3$  change in PM<sub>2.5</sub>.

<sup>b</sup>Correct model includes indicator for estimated date of conception (2-week intervals), gestational age (weeks), maternal tobacco use, education, age, and Medicaid status, and Zip code-specific percent below poverty line and PM<sub>2.5</sub> (see text).

AUC indicates the area under the receiver operating curve; it is 0.5 in the absence of discriminatory ability and 1.0 for perfect discriminatory ability.

**TABLE 2.** Simulation Results

Scenario	Median $\beta^a$ (SE ( $\hat{\beta}$ ))	Percent Bias	Median ( $I$ )	Proportion of Simulations in Which Null Is Rejected ( $H_0: \delta_f = 0$ )	AUC
1 (Correct model) <sup>b</sup>	-12.1 (8.0)	—	-0.01	0.052	0.50
2 (Drop tobacco)	-6.90 (8.0)	44	0.28	0.054	0.51
3 (Drop age)	-8.9 (8.0)	26	0.07	0.048	0.50
4 (Drop percent below poverty line)	-18.4 (8.0)	52	-0.55	0.106	0.55
5 (Drop Medicaid, tobacco, age, education)	7.9 (8.2)	165	1.27	0.224	0.68
6 (Drop race)	-48.2 (8.0)	298	2.00	0.521	0.86

True model (correct) stipulates an effect of exposure (NO<sub>2</sub>) on birth weight.

<sup>a</sup> $\beta$  is the estimated change in birth weight (g), per 20 ppb change in NO<sub>2</sub>.

<sup>b</sup>Correct model includes indicator for estimated date of conception (2-week intervals), gestation age (weeks), maternal tobacco use, education, age, and Medicaid status, and Zip code-specific percent below poverty line and NO<sub>2</sub> (see text).

TABLE 3. Simulation Results

Scenario	Median $\beta^a$ (SE ( $\hat{\beta}$ ))	Percent Bias	Median ( $I$ )	Proportion of Simulations in Which Null Is Rejected ( $H_0: \delta_r = 0$ )	AUC
1 (Correct model) <sup>b</sup>	-20.9 (17.2)	—	-0.01	0.050	—
2 (Drop covariate 1 <sup>c</sup> )	-10.0 (17.4)	3	0.35	0.056	0.52
3 (Drop covariate 2 <sup>c</sup> )	31.2 (18.9)	49	1.59	0.345	0.79
4 (Drop covariate 3 <sup>c</sup> )	78.1 (20.8)	274	2.08	0.690	0.93

True (correct) model stipulates an effect of exposure (PM<sub>2.5</sub>) on birth weight.  
Hypothetical factors Cov1–Cov3 were generated, with specified effects and correlations.  
<sup>a</sup> $\hat{\beta}$  is the estimated change in birth weight (g), per 10  $\mu\text{g}/\text{m}^3$  change in PM<sub>2.5</sub>.  
<sup>b</sup>Correct model includes indicator for estimated date of conception (2-week intervals), gestation age (weeks), maternal tobacco use, education, age, and Medicaid status, and Zip code-specific percent below poverty line, PM<sub>2.5</sub>, and hypothetical variables covariate 1–3, with progressively stronger effect on birth weight (see text and footnote c).  
<sup>c</sup>True  $\beta$  for covariates 1, 2, and 3: 50, 200, and 300, respectively.

factors (Table 3). Although the ability to detect confounding again varied, the indicator consistently signaled possible residual confounding with these sample sizes when the degree of simulated confounding was moderate-to-strong (eg, when race or several variables were omitted [Table 2, scenarios 5 and 6]).

## DISCUSSION

We extend a method to detect important residual confounding<sup>2,3</sup> by describing and evaluating the method for spatial studies. The ability to detect residual confounding was excellent for some scenarios, such as when race was intentionally omitted. As with any statistical technique, the ability to detect residual confounding improves with stronger confounding and larger sample size. We omitted measured variables, such as race, merely to illustrate possible scenarios based on relationships of real factors. In actual applications, the factor creating confounding, if any, could be completely unrecognized and unmeasured. Although few researchers would omit race from a study of air pollution and birth weight, an investigator could conceivably be unaware of, and therefore omit, some other factor that affected air pollution and birth weight in a manner similar to race.

The validity of this approach depends on the assumptions. False-positive indications could arise if, for example, a factor affected both the outcome and future exposures but not the exposure of interest. Our simulations suggest that the method can discriminate situations where residual confounding is present from those where it is not, although the strength of this discrimination ability varies according to the situation.

## REFERENCES

- Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002;155:176–184.
- Flanders WD, Klein M, Strickland M, et al. A method of identifying residual confounding and other violations of model assumptions. *Epidemiology*. 2009;20:S44–S45.
- Flanders WD, Klein M, Strickland M, et al. A method for detection of residual confounding in time-series and other observational studies. *Epidemiology*. 2011;22:59–67.
- Strickland MJ, Klein M, Darrow LA, et al. The issue of confounding in epidemiological studies of ambient air pollution and pregnancy outcomes. *J Epidemiol Community Health*. 2009;63:500–504.
- Greenland S, Pearl J, Robins J. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
- Greenland S, Pearl J. Causal diagrams. In: Boslaugh S, ed. *Encyclopedia of Epidemiology*. Thousand Oaks, CA: Sage Publications; 2007:149–156.
- Ivy D, Mullholland JA, Russell AG. Development of ambient air quality population-weighted metrics for use in time-series health studies. *J Air Waste Manag Assoc*. 2008;58:711–720.