



Practice of Epidemiology

Propensity Score-based Sensitivity Analysis Method for Uncontrolled Confounding

Lingling Li*, Changyu Shen, Ann C. Wu, and Xiaochun Li

* Correspondence to Dr. Lingling Li, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 133 Brookline Avenue, 6th floor, Boston, MA 02215 (e-mail: lingling_li@post.harvard.edu).

Initially submitted January 19, 2010; accepted for publication March 3, 2011.

The authors developed a sensitivity analysis method to address the issue of uncontrolled confounding in observational studies. In this method, the authors use a 1-dimensional function of the propensity score, which they refer to as the sensitivity function (SF), to quantify the hidden bias due to unmeasured confounders. The propensity score is defined as the conditional probability of being treated given the measured covariates. Then the authors construct SF-corrected inverse-probability-weighted estimators to draw inference on the causal treatment effect. This approach allows analysts to conduct a comprehensive sensitivity analysis in a straightforward manner by varying sensitivity assumptions on both the functional form and the coefficients in the 1-dimensional SF. Furthermore, 1-dimensional continuous functions can be well approximated by low-order polynomial structures (e.g., linear, quadratic). Therefore, even if the imposed SF is practically certain to be incorrect, one can still hope to obtain valuable information on treatment effects by conducting a comprehensive sensitivity analysis using polynomial SFs with varying orders and coefficients. The authors demonstrate the new method by implementing it in an asthma study which evaluates the effect of clinician prescription patterns regarding inhaled corticosteroids for children with persistent asthma on selected clinical outcomes.

confounding factors (epidemiology); inverse probability weighting; propensity score; sensitivity analysis; sensitivity function; uncontrolled confounding

Abbreviations: IPW, inverse-probability-weighted; SF, sensitivity function.

Uncontrolled confounding remains a major concern for comparative effectiveness and safety results obtained from analyzing observational studies (1). Sensitivity analysis is of paramount importance and usefulness in assessing the effect of possible uncontrolled confounding on the estimates of the parameter of interest.

Sensitivity analysis for uncontrolled confounding has been studied by multiple researchers. Cornfield et al. (2) first conducted a formal sensitivity analysis examining the association between smoking and lung cancer. Rosenbaum (3) has also done extensive work in sensitivity analysis by modeling the associations between an unobserved confounder and the treatment variable and the outcome of interest. McCandless et al. (4) proposed a Bayesian sensitivity analysis which uses hierarchical prior distributions to infer information on the unobserved confounder using the mea-

sured confounders. Related statistical methods and research are described in detail elsewhere (5–15).

Robins et al. (16) and Brumback et al. (17) proposed an alternative sensitivity analysis method for inverse-probability-weighted (IPW) estimators (18, 19). They quantify the hidden bias due to uncontrolled confounding using a sensitivity function (SF) which depends on the measured potential confounders. In this article, on the basis of their existing work, we propose a new sensitivity analysis approach with a 1-dimensional, propensity score-based SF. The propensity score is defined as the conditional probability of being treated, given measured covariates. We construct the SF-corrected IPW estimators to draw inference on the causal treatment effect. Our new approach is easier and more straightforward to implement. By reducing the dimension of the SF, we make it much easier to specify

the sensitivity functional forms and the values of coefficients. Furthermore, a 1-dimensional continuous function can be reasonably approximated by low-order polynomials (e.g., linear or quadratic) (20, 21). Therefore, even if the imposed SF is practically certain to be incorrect, we can still hope to understand the possible impact of uncontrolled confounding by conducting a comprehensive sensitivity analysis using polynomial SFs with varying orders and coefficients.

CAUSAL INFERENCE IN THE ABSENCE OF UNCONTROLLED CONFOUNDING

Suppose we have n independently and identically distributed copies of data $\{(Y_i, Z_i, X_i), i = 1, \dots, n\}$, where Y_i indicates subject i 's observed outcome, Z_i is the dichotomous treatment variable with 1 for treatment and 0 for control, and X_i is a vector of measured confounders, either continuous or discrete. We also define $Y_{z,i}$ as the potential outcome for treatment $Z_i = z, z \in \{0, 1\}$. Suppose we are interested in estimating the average treatment effect $\psi \equiv E[Y_{1,i}] - E[Y_{0,i}]$. If the outcome is binary, ψ is the causal risk difference. Our results can be easily extended to other causal measures such as risk ratios and odds ratios.

The IPW approach has been well established for deriving causal inference in observational studies in the absence of uncontrolled confounding (18, 19, 22). Its heuristic idea is to construct a pseudopopulation consisting of $w_i = \left[e(X_i)^{Z_i} (1 - e(X_i))^{1-Z_i} \right]^{-1}$ copies of subject i 's data (X_i, Z_i, Y_i) to remove confounding. Here $e(X_i) \equiv \Pr(Z_i = 1|X_i)$ is the so-called propensity score, that is, the conditional probability of being treated given the measured confounders. Specifically, let $\hat{e}(X_i)$ be the estimated propensity score for subject i . Then the IPW estimator of ψ is $\hat{\psi} \equiv \hat{\mu}_1 - \hat{\mu}_0$, where

$$\hat{\mu}_z \equiv \left(\frac{1}{n} \sum_{i=1}^n I(Z_i = z) \hat{w}_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n I(Z_i = z) \hat{w}_i Y_i$$

and

$$\hat{w}_i = \left[\hat{e}(X_i)^{Z_i} (1 - \hat{e}(X_i))^{1-Z_i} \right]^{-1},$$

and $z \in \{0, 1\}$. The idea of IPW originates from survey sampling (23) and has been further generalized to address many issues such as confounding bias in observational studies and missing data (18, 19, 24, 25). When there is no uncontrolled confounding, the potential outcomes are independent of the treatment variable given the values of X , that is, $(Y_0, Y_1) \perp\!\!\!\perp Z|X$, where $\perp\!\!\!\perp$ indicates independence in probability. As a direct consequence, the potential outcomes are also independent of Z given the values of the propensity score, that is, $(Y_0, Y_1) \perp\!\!\!\perp Z|e(X)$. Therefore, it can be shown that $\hat{\psi}$ is a consistent estimator of ψ as long as there is no uncontrolled confounding and $\hat{e}(X_i)$ consistently estimates the true propensity score. However, $\hat{\psi}$ is likely to be biased when X does not include all confounders.

SENSITIVITY ANALYSIS FOR UNCONTROLLED CONFOUNDING

Next, we introduce a new sensitivity analysis approach which is based on the IPW estimators and uses a propensity score-based SF to quantify the hidden bias due to unmeasured confounders.

Sensitivity function

In the presence of uncontrolled confounding, (Y_0, Y_1) is likely to be correlated with Z , conditional on the propensity score $e(X)$. Thus, we cannot obtain a valid estimate of the causal effect by directly comparing the outcome means between the 2 treatment groups in the pseudopopulation. To address this issue, we propose an SF, defined below, to quantify the hidden bias.

$$c(z, e) = E[Y_z|Z = 1, e(X) = e] - E[Y_z|Z = 0, e(X) = e]$$

for $z \in \{0, 1\}$ and $0 < e < 1$. That is, for the subpopulation with the same propensity score values (i.e., $e(X) = e$), $c(z, e)$ is the mean difference for the potential outcome Y_z between the treated ($Z = 1$) and untreated ($Z = 0$) groups. In other words, $c(z, e)$ quantitatively measures the impact of unmeasured confounders on the difference in potential outcomes between the treated and untreated subjects, conditional on the propensity score of measured covariates. The range of the SF $c(z, e)$ reflects the magnitude of uncontrolled confounding bias. Under the assumption of no uncontrolled confounding, $c(z, e) = 0$ for any value of (z, e) .

Our sensitivity analysis method is developed along the lines of the work by Robins et al. (16, 26) and Brumback et al. (17), in which the uncontrolled confounding is quantified using the between-group difference, conditional on the values of all measured confounders, that is, $c^*(z, x) \equiv E[Y_z|Z = 1, X = x] - E[Y_z|Z = 0, X = x]$. We prove in the Web Appendix (available on the *Journal's* Web site (<http://aje.oxfordjournals.org/>)) that $c(z, e) = E[c^*(z, X)|e(X) = e]$. Therefore, if $c^*(z, X)$ is constant, $c^*(z, X) = c(z, e)$, and both methods are the same. Nonetheless, when X contains multiple covariates, our approach will be much easier to implement because our SF depends only on a single random variable $e(X)$, a 1-dimensional summary of X , rather than a multidimensional vector X . Note that in performing a sensitivity analysis, an analyst needs to specify not only the functional form of the SF but also the values of the coefficients. For instance, $c^*(z, x)$ equals $\beta_1 \times \text{age} + \beta_2 \times \text{race}$, where β_1 equals 0.2 and β_2 equals 0.1. When $c^*(z, x)$ is expected to depend on multiple confounders, such specifications will be difficult, and the imposed working functional form is unlikely to accurately reflect the complex relations between the measured and unmeasured confounders and the potential outcomes. Furthermore, since we cannot empirically verify the imposed assumptions using the observed data, it is a common practice to vary these assumptions in sensitivity analysis to evaluate the corresponding causal estimates. When $c^*(z, x)$ is high-dimensional, it will be technically difficult to do so, as we would need to vary many parameters in $c^*(z, x)$ simultaneously.

Our approach nicely reduces the dimension of the SF and makes it much easier to vary sensitivity assumptions to explore plausible scenarios. In practice, the specified SF is likely to be incorrect. Nonetheless, since the new SF is 1-dimensional, low-order polynomials (e.g., linear, quadratic) should be able to provide reasonable approximations as long as the true SF $c(z, e)$ is continuous in the interval $[0, 1]$ (20, 21). We suggest conducting a sensitivity analysis with constant, linear, or quadratic SFs with the coefficients varying over a set of plausible values, which should be selected on the basis of the observed data, literature, and subject knowledge specific to the application setting. For instance, suppose the outcome of interest is death and the treated subjects are relatively sicker than the untreated subjects; then, before conducting sensitivity analysis, we need to understand how different the treated and untreated groups are after controlling for measured confounders. Suppose we expect an average 5%–10% excess risk for the treated subjects as compared with the untreated subjects, even if, contrary to fact, they were given the same treatment. We would vary $c(z, e)$ between 0.05 and 0.1 when considering a constant SF. If we expect the amount of hidden bias to vary approximately linearly across levels of propensity score, we could use a linear SF. Since $e \in [0, 1]$, we would select the intercept and slope of the linear SF on the basis of the likely values of excess risks for persons with very large or very small propensity scores. For instance, $c(z, e) = 0.05 + 0.05e$ if we expect $c(z, e)$ to increase with e or $c(z, e) = 0.1 - 0.05e$ if we expect $c(z, e)$ to decrease with e . Later in this article, we provide more specific illustrations and instructions in the context of an example.

SF-corrected IPW estimators

Given an SF $c(z, e)$, we construct the SF-corrected IPW estimators by replacing the observed Y in the original estimator $\hat{\psi}$ with the SF-corrected outcomes

$$\begin{aligned} Y^{\text{SF}} &\equiv Y - \left(E[Y_Z|Z, e(X)] - E[Y_Z|e(X)] \right) \\ &= \begin{cases} Y - (1 - e(X))c(1, e) & \text{if } Z = 1, \\ Y + e(X)c(0, e) & \text{if } Z = 0. \end{cases} \end{aligned} \quad (1)$$

Then,

$$\hat{\psi}^{\text{SF}} = \hat{\mu}_1^{\text{SF}} - \hat{\mu}_0^{\text{SF}}, \quad (2)$$

where

$$\hat{\mu}_z^{\text{SF}} = \left(\frac{1}{n} \sum_{i=1}^n I(Z_i = z) w_i \right)^{-1} \frac{1}{n} \sum_{i=1}^n I(Z_i = z) w_i Y_i^{\text{SF}}.$$

We prove in the Web Appendix the consistency of $\hat{\psi}^{\text{SF}}$. The intuitive idea is straightforward: We simply remove the hidden bias in the pseudopopulation by applying the SF. Then the causal effect can be consistently estimated by the between-group difference of outcome means.

Note that for noncontinuous outcomes (e.g., binary outcomes), instead of using an additive SF as defined above, we may use a multiplicative SF

$$c(z, e) \equiv \log \frac{E[Y_Z|Z = 1, e]}{E[Y_Z|Z = 0, e]}$$

and remove residual bias by defining

$$Y^{\text{SF}} \equiv Y \frac{E[Y_Z|e(X)]}{E[Y_Z|Z, e(X)]},$$

which equals $Y\{e(X) + \exp(-c(z = 1, e(X))(1 - e(X)))\}$ if $Z = 1$ and $Y\{1 - e(X) + \exp(c(z = 0, e(X))e(X))\}$ if $Z = 0$. Then it can be easily shown that $E[Y^{\text{SF}}|Z = z, X] = E[Y_Z|e(X)]$. In addition, Y^{SF} is guaranteed to be positive. Nonetheless, with binary outcomes, the estimates of the marginal means $\hat{\mu}_1^{\text{SF}}$ and $\hat{\mu}_0^{\text{SF}}$ may still be outside the plausible range $[0, 1]$. This in turn guides analysts in the selection of sensible parameters in the SF.

In this article, we chose to focus primarily on the additive SF for the purposes of illustration and explication. Multiplicative SFs can be implemented in exactly the same manner.

Variance and bootstrap confidence intervals

We estimate the propensity score using generalized boosted models (27). The generalized boosted models approach predicts treatment assignment from a large number of pretreatment covariates through adaptive choice of variables. It inherits many of the properties of regression trees, and thus is flexible, can capture complex interactions among confounders and nonlinear terms, and can handle highly correlated covariates. Ridgeway and McCaffery (28) have provided empirical evidence, via a simulation study, showing that the generalized boosted models method produces more stable and reliable estimates of the propensity score than logistic regression models. We then use bootstrap methods (29) to obtain estimates of the variances and the 95% bootstrap confidence intervals. Specifically, the variances are estimated with the sample variances and the bootstrap confidence intervals are estimated with the corresponding percentiles among the bootstrap realizations.

APPLICATION TO THE ASTHMA STUDY

To illustrate our method, we present data from a Boston, Massachusetts, prospective cohort study (the Parent Asthma Communication Experience Study) comparing selected clinical outcomes during a 12-month follow-up period (2007–2008) for children with mild persistent asthma whose parents believed that they were supposed to use inhaled corticosteroids either daily or periodically (30).

Motivating example

Study subjects with mild persistent asthma were initially identified using administrative claims records. A diagnosis of persistent asthma was later confirmed on the basis of responses to a telephone survey (30). Our study cohort

Table 1. Distribution of Children With Mild Persistent Asthma and Risks of Uncontrolled Asthma According to Treatment Group (Daily Use or Periodic Use of Inhaled Corticosteroids) and Associated Risk Differences, Parent Asthma Communication Experience Study, 2007–2008

PS Stratum ^a	PS Cutpoint	Daily Use		Periodic Use		Risk Difference
		No.	Risk	No.	Risk	
1	0.30	7	0.14	66	0.20	-0.06
2	0.46	22	0.36	50	0.32	0.04
3	0.62	37	0.49	36	0.36	0.13
4	0.77	57	0.46	15	0.20	0.26
5	0.96	71	0.66	2	0.00	0.66
Total		194	0.52	169	0.27	0.25

Abbreviation: PS, propensity score.

^a Estimated quintiles.

consisted of 363 children with a confirmed diagnosis of mild persistent asthma. The children were classified into 2 treatment groups based on whether their parents believed that their health-care providers had told them to have their children use inhaled corticosteroids daily ($Z = 1$) or periodically (i.e., as needed; $Z = 0$). The clinical outcome of interest (Y) was whether a child had an episode of uncontrolled asthma during the 12-month period after the telephone survey. The measured confounders X included age, race, parental education, household income, health insurance provider, baseline asthma physical status score, competing family priorities score, and health-care utilization during the baseline period (i.e., 12 months before the survey). Crude comparisons showed that the periodic users had better observed outcomes than the daily users. We hypothesized that periodic users appeared to have fewer episodes of uncontrolled asthma because the periodic users had less severe asthma at baseline. Our goal was to evaluate whether periodic users had worse outcomes than daily users, while adjusting for both measured covariates and possibly uncontrolled confounding. The causal effect ψ indicated the effect of the pattern of inhaled corticosteroid prescription on the clinical outcome of uncontrolled asthma—that is, the rate difference for uncontrolled asthma supposing that the entire study population was told to use inhaled corticosteroids daily versus periodically. We did not intend to evaluate the treatment effect of inhaled corticosteroids because a new-user design (31) was infeasible and we were not able to monitor actual use of the medications.

Preliminary analyses showed that the daily-use group had a worse baseline asthma physical status score and a higher baseline medication concerns score than the periodic-use group (30). In this specific example, sicker participants were more likely to be treated (i.e., to be prescribed daily use of inhaled corticosteroids). The crude risk difference for uncontrolled asthma was 0.25 (95% confidence interval: 0.15, 0.35), and the adjusted risk difference was 0.20 (95% confidence interval: 0.12, 0.30). In Table 1, we present the numbers of subjects and absolute risks by treatment group, as well as risk differences, in the entire study cohort and

within each propensity score stratum, defined by estimated quintiles. After adjustment for the estimated propensity scores, the marginal distributions for the measured baseline covariates were balanced between the 2 treatment groups without significant differences. In the propensity score estimation, we adjusted for age, race, health maintenance organization site, parental education, household income, baseline asthma physical status score, competing family priorities score, expectations score, provider interaction score, number of outpatient visits for asthma during the baseline period (i.e., the 12 months before the interview), uncontrolled asthma during the baseline period, and the number of emergency department visits or hospitalizations in the baseline period.

The results imply that the daily-use pattern is worse than the periodic-use pattern, especially among subjects with high propensity scores. These results are inconsistent with both the anecdotal clinical experience of physicians and national guidelines which recommend that children with persistent asthma use inhaled corticosteroids on a daily basis year-round. We suspect that the paradoxical results are due to uncontrolled confounding. The seemingly harmful effect of daily use is probably due to the fact that the children who were prescribed daily use of inhaled corticosteroids had more severe disease during the baseline period, which was not adequately captured by the measured covariates. On the basis of the observed increasing risk difference from the lowest propensity score stratum to the highest stratum, we hypothesize that the magnitude of residual bias due to unmeasured confounders increases with the estimated propensity score; that is, subjects with high propensity scores are probably children with the most severe asthma and are thus subject to more uncontrolled confounding bias. Our sensitivity analyses will examine whether this harmful effect of daily use diminishes as we account for possible residual confounding bias.

Note that Table 1 shows some poor overlap of the propensity score distributions between the 2 treatment groups, as there are very few daily users in stratum 1 and very few periodic users in stratum 5. In this article, we use this example purely for the purpose of illustrating the sensitivity analysis method and thus do not wish to distract the reader with extra details. In real-life applications, analysts may consider redefining the study population such that their estimated propensity scores are on the overlapping support (32).

Sensitivity analysis

In the sensitivity analysis, we consider constant, linear, and quadratic SFs. In all analyses, we use the generalized boosted models approach to estimate the propensity score (33). The 95% bootstrap confidence intervals are constructed using the 2.5% and 97.5% percentiles of the 1,000 bootstrap realizations of $\hat{\psi}^{\text{SF}}$.

Constant SF. We first consider $c(z, e) = c_z$ for $z \in \{0, 1\}$. In our motivating example, we expect the treated group (daily-use pattern) to have poorer pretreatment asthma status than the untreated group (periodic-use pattern), even after controlling for measured confounders. Thus, we expect c_0 and c_1 to be nonnegative constants, since children with

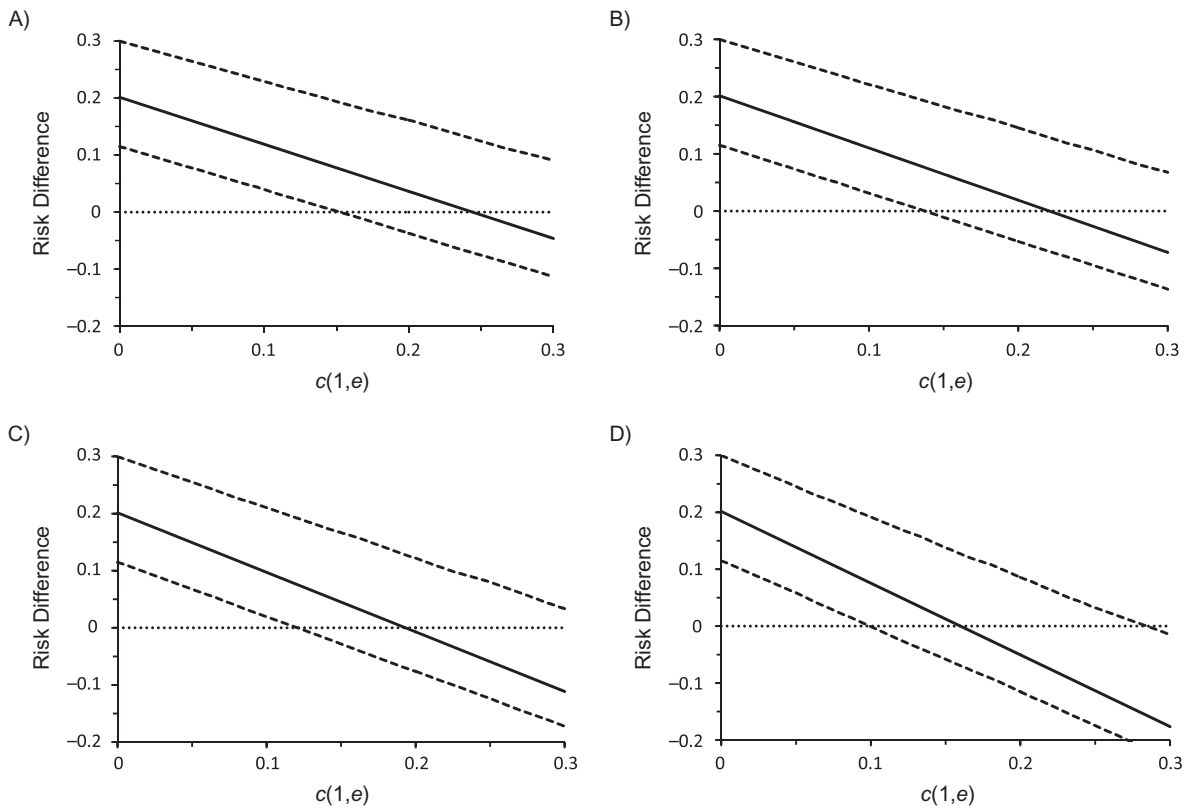


Figure 1. Point estimates (solid line) of the risk difference for uncontrolled asthma between daily use and periodic use of inhaled corticosteroids in children with mild persistent asthma, assuming constant sensitivity functions, Parent Asthma Communication Experience Study, 2007–2008. The authors vary the values of $c(1, e)$ and $r = c(0, e)/c(1, e)$. A) $r = 1$; B) $r = 1.2$; C) $r = 1.5$; D) $r = 2$. The 95% bootstrap confidence intervals (dotted lines) were obtained using the 2.5% and 97.5% percentiles among the 1,000 bootstrap replications.

poorer asthma status are likely to have higher risks of uncontrolled asthma during the follow-up period than the other group, even if, contrary to fact, they received the same treatment. Note that disease scores typically measure a different dimension than propensity scores. Our sensitivity assumptions in this asthma example are made on the basis of observed data, literature, and subject knowledge for this specific application, and they may not apply to other settings. In the crude analysis prior to adjustment for any confounders, 52% of treated subjects and 27% of untreated subjects had uncontrolled asthma. Based on clinical knowledge and experience, we do not expect the daily-use pattern to produce worse clinical outcomes than the periodic-use pattern; therefore, we vary the value of c_1 in the range of $[0, 0.3]$. We allow for the possibility that c_0 is greater than c_1 —that is, that the uncontrolled confounding has a bigger impact on the potential outcome for periodic use (Y_0) than on the potential outcome for daily use (Y_1).

Let $r \equiv c_0/c_1$ indicate the ratio between the 2 constant SFs. In Figure 1, we present 4 plots for r values of 1, 1.2, 1.5, and 2, respectively. The solid lines indicate the point estimates of the risk difference, while the dotted lines indicate the lower and upper limits of the 95% bootstrap confidence intervals. The horizontal line represents the null value of 0.

As expected, the risk difference estimates decrease when either c_1 or r increases. This is intuitively plausible, because the more the uncontrolled confounding is assumed to exist, the further the SF-corrected risk difference estimator decreases as we attribute an increasing proportion of the observed risk difference to the effect of uncontrolled confounding. Under the assumption of no unmeasured confounder, the estimated risk difference for uncontrolled asthma is 0.2 (95% bootstrap confidence interval: 0.1, 0.3); that is, the daily-use pattern leads to a 20% excess risk. Let us first examine the plot in the upper left corner of Figure 1 with $r = 1$ (i.e., $c(z = 1, e) = c(z = 0, e) = c_1$). When c_1 increases to approximately 0.15 (i.e., the treated group has a 0.15 greater risk of uncontrolled asthma than the uncontrolled group, regardless of whether they were all treated or all untreated), the lower bound of the 95% bootstrap confidence interval crosses the null value of zero, indicating an insignificant risk difference between the 2 patterns of inhaled corticosteroid use. The 95% bootstrap confidence interval remains statistically insignificant within the considered range of $c_1 \leq 0.3$. The point estimate for the risk difference decreases to -0.05 at $c_1 = 0.3$, but this difference is insignificant. Results shown in other parts of Figure 1 with varying values of r are very similar. The 95%

Table 2. Estimated Risk Difference for Uncontrolled Asthma According to Treatment Group (Daily Use or Periodic Use of Inhaled Corticosteroids) Among Children With Mild Persistent Asthma, Parent Asthma Communication Experience Study, 2007–2008^a

c_1	s_1	$r = 1.0$		$r = 1.5$		$r = 2.0$	
		RD	95% BCI ^b	RD	95% BCI	RD	95% BCI
0.00	0.00	0.20	0.12, 0.30	0.20	0.12, 0.30	0.20	0.12, 0.30
	0.05	0.18	0.10, 0.28	0.18	0.10, 0.28	0.17	0.09, 0.28
	0.10	0.16	0.08, 0.27	0.15	0.07, 0.26	0.14	0.06, 0.25
	0.20	0.12	0.05, 0.23	0.10	0.03, 0.22	0.08	0.01, 0.20
	0.30	0.08	0.01, 0.20	0.05	-0.02, 0.18	0.02	-0.05, 0.15
0.05	0.00	0.16	0.08, 0.27	0.15	0.07, 0.26	0.14	0.07, 0.25
	0.05	0.14	0.06, 0.25	0.13	0.05, 0.24	0.11	0.04, 0.22
	0.15	0.10	0.03, 0.22	0.08	0.00, 0.20	0.05	-0.02, 0.17
	0.25	0.06	-0.01, 0.18	0.03	-0.05, 0.15	-0.01	-0.08, 0.12
0.10	0.00	0.12	0.05, 0.23	0.10	0.03, 0.22	0.08	0.01, 0.20
	0.10	0.08	0.01, 0.20	0.05	-0.02, 0.17	0.02	-0.05, 0.15
	0.20	0.04	-0.03, 0.16	0.00	-0.07, 0.13	-0.04	-0.11, 0.10
0.20	0.00	0.04	-0.03, 0.16	0.00	-0.07, 0.13	-0.04	-0.11, 0.09
	0.10	0.00	-0.07, 0.13	-0.05	-0.12, 0.09	-0.10	-0.17, 0.04
0.30	0.00	-0.05	-0.10, 0.10	-0.10	-0.16, 0.05	-0.16	-0.23, -0.01

Abbreviations: BCI, bootstrap confidence interval; RD, risk difference.

^a The authors assumed that the sensitivity functions followed linear structures, such that $c(z = 1, e) = c_1 + s_1 e$ and $c(z = 0, e) = r \times (z = 1, e)$.

^b The 95% bootstrap confidence intervals were obtained using the 2.5% and 97.5% percentiles among the 1,000 bootstrap replications.

bootstrap confidence interval becomes significant only in the unlikely setting in which c_1 is approximately 0.3 and c_0 is approximately 0.6.

Linear and quadratic SFs. We also consider the linear SF $c(z, e) = c_z + s_z e(X)$; that is, the effect of uncontrolled confounding changes linearly with the propensity score. In this example, we expect the magnitude of residual confounding to increase with the propensity score (i.e., s_z is positive), since children with higher propensity scores are expected to be sicker at baseline. Then c_z and $c_z + s_z$ indicate the lower and upper bounds of $c(z, e)$, respectively. Note that the propensity score is likely to be bounded away from 0 and 1. However, for this specific example, we do not have enough background information to specify the exact boundaries.

In Table 2, we present the results for a set of scenarios in which c_1 varies between 0 and 0.3 and s_1 varies between 0 and $0.3 - c_1$ for a given c_1 , since we do not expect $c(1, e)$ to exceed 0.3. We also consider 3 possible values (i.e., 1.0, 1.5, and 2.0) for the ratio $r = c(0, e)/c(1, e)$. The numbers presented in Table 2 suggest that the results are similar to those shown in Figure 1. The point and interval estimates of the risk difference keep decreasing when c_1 , s_1 , or r increases. The more the uncontrolled confounding is assumed to exist, the smaller the risk difference estimates are. With larger values of r , the estimates decrease at an even faster rate. Under certain scenarios, the point estimates of risk difference are negative, indicating a protective effect of daily use on the incidence of uncontrolled asthma during

the follow-up period. However, the differences remain statistically insignificant under plausible scenarios.

To evaluate the effect of violations of linear structures, we further consider the quadratic SF $c(z, e) = c_z + s_z e + q_z e^2$. The coefficients (c_z , s_z , q_z) do not have direct interpretations. Their values are determined by the values of the SF at 3 points—for example, $c(z, e = 0)$ (the lower bound), $c(z, e = 1)$ (the upper bound), and $c(z, e = 0.5)$. For each selected ($c(z, 0)$, $c(z, 1)$), we vary the value of $c(z, e = 0.5)$ between $c(1, e = 0) + (1/4)[c(1, e = 1) - c(1, e = 0)]$ and $c(1, e = 0) + (3/4)[c(1, e = 1) - c(1, e = 0)]$. The difference between the middle point $c(z, e = 0.5)$ and $(1/2)[c(z, e = 0) + c(z, e = 1)]$ indicates the deviation of the SF from linear structures. The results are very similar and thus are not shown here.

In summary, we conducted a comprehensive sensitivity analysis for the asthma study considering constant, linear, and quadratic SFs and various sets of coefficients. After accounting for possible uncontrolled confounding, the unlikely harmful effect of the daily-use pattern diminishes. The risk difference estimates become negative, suggesting some beneficial effect of the daily-use prescription pattern when $c(z, e)$ is 0.25 or higher. Nevertheless, the differences are statistically insignificant under a wide range of plausible scenarios. Thus, our study population does not exhibit strong evidence supporting the superiority of the daily-use pattern compared with the periodic-use pattern. The results are consistent with some providers' clinical experience that periodic use of inhaled corticosteroids is effective for

selected patients with mild intermittent asthma. The 2007 National Heart, Lung, and Blood Institute guidelines (34) also state that periodic or seasonal treatment is an acceptable option for some children. Several recent clinical trials have demonstrated the effectiveness of periodic inhaled corticosteroid use for selected adult patients with mild persistent asthma (35–37). However, to our knowledge, no studies exist for children. The published study on this asthma example (30) provides useful information with which to guide clinical practice for children with asthma. Nonetheless, the results suffer from the bias due to unmeasured confounders. Our sensitivity analysis directly addresses this issue and provides a comprehensive assessment of the 2 use patterns to answer the critical clinical question of interest.

DISCUSSION

We have introduced a new propensity score-based sensitivity analysis method that uses the SF-corrected IPW estimators to assess the effect of possible uncontrolled confounding in observational studies. As we have shown through its application to an asthma study (30), the new method can be easily adopted to provide valuable insight on the impact of uncontrolled confounding. The SF is a 1-dimensional function of the propensity score. If strong prior information is available, appropriate functional forms and coefficients can be directly imposed. Otherwise, low-order (e.g., linear, quadratic) polynomials are expected to provide reasonably good approximations of continuous 1-dimensional functions. We suggest varying the coefficients over a set of plausible values, which should be determined on the basis of observed data, literature, and subject knowledge. The sensitivity assumptions we made in the analysis were appropriate for our asthma study but may not apply to other settings. These assumptions need to be examined or modified before our method is applied to other studies.

The proposed method is a direct extension of an existing sensitivity analysis method (17) in which the SF depends on the entire covariate vector X . The motivation is to reduce the dimension of the SF to facilitate the implementation of a comprehensive sensitivity analysis. Nonetheless, a good understanding of the relation between the propensity score and the disease risk is still required in order to impose reasonable parametric assumptions on the 1-dimensional SF. In some settings, patients who have similar propensity scores may have totally different disease risks and thus are subject to different amounts of hidden bias (e.g., relatively healthy patients and very severely impaired patients may both have low propensities of receiving the treatment). Then it would be less straightforward to impose assumptions on the 1-dimensional SF, since we collapse subjects who have similar propensity scores but different disease risks together. In such settings, we suggest taking an intermediate step to balance the trade-off between reducing the dimension of the SF and keeping subjects with different disease risks separate. Specifically, we could define the SF as the conditional mean difference in the potential outcomes between the treated and untreated subgroups, conditional on not only the propensity score but also 1 or several elements in X that

were strong predictors of the outcome (e.g., a dummy variable indicating whether the patient was severely impaired). It can be shown that the aforementioned results apply.

The idea of bias correction using a propensity score-based SF can also be applied to other causal inference methods—for example, propensity score matching and stratification and doubly robust estimation (38). These methods share the same heuristic ideas as the IPW approach in that they use the propensity score to select comparable real-life treated and untreated subjects from whom to borrow information and draw causal inferences. Therefore, the idea of removing the hidden bias using the imposed SF applies equally to all of them. Within each level of the propensity score, the expectation of the SF-corrected outcome in a single treatment group equals the expectation of the corresponding potential outcome in both treatment groups. Then the causal treatment effect can be consistently estimated given that the imposed SF is correct. In future work, we plan to evaluate and compare the performance of different SF-corrected estimators under various scenarios. We will also extend the work to longitudinal settings with repeated measurements.

Finally, note that the proposed sensitivity analysis does not work when the IPW approach does not work. For instance, when the weights are highly variable (propensity scores for some treated subjects are close to 0 and/or propensity scores for some untreated subjects are close to 1), IPW estimators are known to be sensitive and unstable (39). In such settings, our sensitivity analysis approach is unlikely to yield useful information, since it is based on the IPW estimation. Other sensitivity analysis approaches might be preferred. In their 2004 article, Brumback et al. (17) provide an excellent discussion on relevant approaches. However, a major advantage of Brumback's and our approaches is that they can be used to explore sensitivity to multiple unmeasured confounders simultaneously.

To help researchers implement the proposed work, we have written a user-friendly software program within the R statistical computing environment (R Foundation for Statistical Computing, Vienna, Austria) with which to conduct the proposed sensitivity analysis for a list of sensitivity functional forms. The R program will be posted on our faculty Web sites.

ACKNOWLEDGMENTS

Author affiliations: Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts (Lingling Li, Ann C. Wu); Division of Biostatistics, School of Medicine, Indiana University, Indianapolis, Indiana (Changyu Shen, Xiaochun Li); and Division of General Pediatrics, Children's Hospital Boston, Boston, Massachusetts (Ann C. Wu).

Data from the Parent Asthma Communication Experience Study were collected with financial support from a Midcareer Investigator Award in Patient-Oriented Research from the National Institute of Child Health and Human Development (grant K24 HD047667 to Professor Tracy A. Lieu).

The authors are grateful to the investigators and staff of the Parent Asthma Communication Experience Study for collecting the interview data. The authors are especially grateful to Professor Tracy A. Lieu for giving them permission to use the data for development and illustration of their methods.

Preliminary results of this work were presented at the 16th Annual HMO Research Network Conference in Austin, Texas, March 21–24, 2010, and at the 2010 ICSA Applied Statistics Symposium in Indianapolis, Indiana, June 20–23, 2010.

Conflict of interest: none declared.

REFERENCES

- Delaney JA, Platt RW, Suissa S. The impact of unmeasured baseline effect modification on estimates from an inverse probability of treatment weighted logistic model. *Eur J Epidemiol.* 2009;24(7):343–349.
- Cornfield J, Haenszel W, Hammond EC, et al. Smoking and lung cancer: recent evidence and a discussion of some questions. *J Natl Cancer Inst.* 1959;22(1):173–203.
- Rosenbaum P. *Observational Studies.* New York, NY: Springer-Verlag New York; 2002.
- McCandless LC, Gustafson P, Levy AR. A sensitivity analysis using information about measured confounders yielded improved uncertainty assessments for unmeasured confounding. *J Clin Epidemiol.* 2008;61(3):247–255.
- Schlesselman JJ. Assessing effects of confounding variables. *Am J Epidemiol.* 1978;108(1):3–8.
- Greenland S, Neutra R. An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. *J Chronic Dis.* 1981;34(9-10):433–438.
- Flanders WD, Khoury MJ. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology.* 1990;1(3):239–246.
- Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics.* 1998;54(3):948–963.
- Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. *J Am Stat Assoc.* 2003;98(461):47–54.
- Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *Am J Epidemiol.* 2004;160(4):384–392.
- Greenland S. Multiple-bias modelling for analysis of observational data. *J R Stat Soc Ser A Stat Soc.* 2005;168(2):267–306.
- MacLehose RF, Kaufman S, Kaufman JS, et al. Bounding causal effects under uncontrolled confounding using counterfactuals. *Epidemiology.* 2005;16(4):548–555.
- Stürmer T, Glynn RJ, Rothman KJ, et al. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. *Med Care.* 2007;45(10 suppl 2):S158–S165.
- Stürmer T, Schneeweiss S, Rothman KJ, et al. Performance of propensity score calibration—a simulation study. *Am J Epidemiol.* 2007;165(10):1110–1118.
- Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol.* 2008;18(8):637–646.
- Robins JM, Rotnitzky A, Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran ME, Berry D, eds. *Statistical Models in Epidemiology: The Environment and Clinical Trials.* New York, NY: Springer-Verlag New York; 1999:1–92.
- Brumback BA, Hernán MA, Haneuse SJ, et al. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med.* 2004;23(5):749–767.
- Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000;11(5):561–570.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000;11(5):550–560.
- Rudin W. *Functional Analysis.* New York, NY: McGraw-Hill Company; 1973.
- Rudin W. *Principles of Mathematical Analysis.* Vol 3. New York, NY: McGraw-Hill Company; 1976.
- Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran E, Berry D, eds. *Statistical Models in Epidemiology: The Environment and Clinical Trials.* New York, NY: Springer-Verlag New York; 1999:95–134.
- Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc.* 1952;47(260):663–685.
- Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc.* 1994;89(427):846–866.
- Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J Am Stat Assoc.* 1995;90(429):106–121.
- Robins JM. Association, causation, and marginal structural models. *Synthese.* 1999;121(1-2):151–179.
- McCaffrey DF, Ridgeway G, Morral AR. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychol Methods.* 2004;9(4):403–425.
- Tsiatis AA, Davidian M. Comment: demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Stat Sci.* 2007;22(4):569–573.
- Efron B, Tibshirani R. *An Introduction to the Bootstrap.* Boca Raton, FL: CRC Press; 1997.
- Wu AC, Li L, Miroshnik I, et al. Outcomes after periodic use of inhaled corticosteroids in children. *J Asthma.* 2009;46(5):517–522.
- Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol.* 2003;158(9):915–920.
- Schneeweiss S, Patrick AR, Solomon DH, et al. Variation in the risk of suicide attempts and completed suicides by antidepressant agent in adults: a propensity score-adjusted analysis of 9 years' data. *Arch Gen Psychiatry.* 2010;67(5):497–506.
- Comprehensive R Archive Network. *twang: Toolkit for Weighting and Analysis of Nonequivalent Groups* [software]. Vienna, Austria: R Foundation for Statistical Computing; 2009. (<http://cran.r-project.org/web/packages/twang/index.html>). (Accessed January 15, 2010).
- National Heart, Lung, and Blood Institute. *Guidelines for the Diagnosis and Management of Asthma (Expert Panel Report 3).*

- Bethesda, MD: National Heart, Lung, and Blood Institute; 2007. (<http://www.nhlbi.nih.gov/guidelines/asthma/>). (Accessed January 15, 2010).
35. Boushey HA, Sorkness CA, King TS, et al. Daily versus as-needed corticosteroids for mild persistent asthma. National Heart, Lung, and Blood Institute's Asthma Clinical Research Network. *N Engl J Med.* 2005;352(15):1519–1528.
 36. Papi A, Canonica GW, Maestrelli P, et al. Rescue use of beclomethasone and albuterol in a single inhaler for mild asthma. BEST Study Group. *N Engl J Med.* 2007;356(20):2040–2052.
 37. Peters SP, Anthonisen N, Castro M, et al. Randomized comparison of strategies for reducing treatment in mild persistent asthma. American Lung Association Asthma Clinical Research Centers. *N Engl J Med.* 2007;356(20):2027–2039.
 38. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics.* 2005;61(4):962–973.
 39. Kurth T, Walker AM, Glynn RJ, et al. Results of multi-variable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol.* 2006;163(3):262–270.