Appendix

Proof of Theorem 1. We give the proof for d_{a_+}. The proofs for d_{a_1} and d_{a_0} are given in the eAppendix (http://links.lww.com).

$$\begin{aligned}
d_{a_+} &= \{\textstyle\sum_x E(Y|a_1,x)P(x) - \sum_x E(Y|a_0,x)P(x)\} - \{E(Y_{a_1}) - E(Y_{a_0})\} \\
&= \textstyle\sum_x \sum_u E(Y|a_1,x,u)P(u|a_1,x)P(x) - \sum_x \sum_u E(Y|a_0,x,u)P(u|a_0,x)P(x) \\
&\qquad - \textstyle\sum_x \sum_u E(Y_{a_1}|x,u)P(u|x)P(x) + \sum_x \sum_u E(Y_{a_0}|x,u)P(u|x)P(x) \\
&\qquad\qquad \text{by iterated expectations} \\
&= \textstyle\sum_x \sum_u E(Y|a_1,x,u)P(u|a_1,x)P(x) - \sum_x \sum_u E(Y|a_0,x,u)P(u|a_0,x)P(x) \\
&\qquad - \textstyle\sum_x \sum_u E(Y_{a_1}|a_1,x,u)P(u|x)P(x) + \sum_x \sum_u E(Y_{a_0}|a_0,x,u)P(u|x)P(x) \\
&\qquad\qquad \text{by unconfoundedness conditional on } (U,X) \\
&= \textstyle\sum_x \sum_u E(Y|a_1,x,u)\{P(u|a_1,x) - P(u|x)\}P(x) \\
&\quad - \textstyle\sum_x \sum_u E(Y|a_0,x,u)\{P(u|a_0,x) - P(u|x)\}P(x) \\
&\qquad\qquad \text{by consistency} \\
&= \textstyle\sum_x \sum_u \{E(Y|a_1,x,u) - E(Y|a_1,x,u')\}\{P(u|a_1,x) - P(u|x)\}P(x) \\
&\qquad - \textstyle\sum_x \sum_u \{E(Y|a_0,x,u) - E(Y|a_0,x,u')\}\{P(u|a_0,x) - P(u|x)\}P(x) \\
&\qquad\qquad \text{since } E(Y|a_1,x,u') \text{ and } E(Y|a_1,x,u') \text{ are constants.} \blacksquare
\end{aligned}$$

Proof of Theorems 2 and 3. See eAppendix (http://links.lww.com) for details.

Relation to Other Sensitivity Analysis Techniques

Relation to the Sensitivity Analysis of Rosenbaum and Rubin (1983)

In the sensitivity analysis proposed by Rosenbaum and Rubin,[6] they consider a binary outcome Y, binary treatment A, covariate(s) X (in their application, X indicates propensity score strata) and hypothesize a binary unmeasured confounder U such that $Y_a \coprod A|X,U$.

The researcher specifies sensitivity parameters

$$\begin{aligned}
\pi_x &= P(U=1|x) \\
\alpha_x &= \log\Big(\frac{P(A=0|U=1,X=x)}{1-P(A=0|U=1,X=x)} / \frac{P(A=0|U=0,X=x)}{1-P(A=0|U=0,X=x)}\Big) \\
\delta_{xt} &= \log\Big(\frac{P(Y=0|U=1,A=t,X=x)}{1-P(Y=0|U=1,A=t,X=x)} / \frac{P(Y=0|U=0,A=t,X=x)}{1-P(Y=0|U=0,A=t,X=x)}\Big).
\end{aligned}$$

For specified values of π_x, α_x and δ_{xt}, maximum likelihood estimates of the causal effect can then be obtained; sensitivity analysis proceeds by specifying different values of the parameters π_x, α_x and δ_{xt}. In their application (described above), π_x, α_x and δ_{xt} are assumed to be constant over x. As noted above, assuming no unmeasured confounding the estimate of the causal effect is 0.31. Rosenbaum and Rubin[6] consider values for π (the overall prevalence of U) of 0.1, 0.5 and 0.9. They first consider $\alpha = 2$ and values $\delta_{t=0}$ and $\delta_{t=1}$ either $\frac{1}{2}$ or 2; under this set of scenarios, the smallest causal effect estimate, an estimate of 0.28, is obtained when the prevalence of U is 0.5 and when U doubles the odds of surgery ($\alpha = 2$) and also doubles the odds of improvement ($\delta_{t=0} = 2$ and $\delta_{t=1} = 2$). They then consider $\alpha = 3$ and values $\delta_{t=0}$ and $\delta_{t=1}$ either $\frac{1}{3}$ or 3; under this set of scenarios, the smallest causal effect estimate, an estimate of 0.25, is obtained when the prevalence of U is 0.5 and when U triples the odds of surgery ($\alpha = 3$) and also triples the odds of improvement ($\delta_{t=0} = 3$ and $\delta_{t=1} = 3$). They conclude that for an unobserved confounder to explain the outcome difference comparing medical and surgical patients, it would have to more than triple the odds of surgery and of improvement. Admittedly, this seems unlikely. We saw above, however, that when the sensitivity analysis is conducted on a risk-difference scale rather than an odds-ratio scale, although the degree of uncontrolled confounding that would be needed to explain away the estimate of the causal effect is still unlikely, the numbers are perhaps slightly less inconceivable.

The bias formula for d_{a_+} in Theorem 1 can also be used in a reasonably straightforward way to replicate the odds-ratio sensitivity analysis of Rosenbaum and Rubin.[6] Let $a_0 = 0$

and $a_1 = 1$ and let $u = 1$ and $u' = 0$. For specified values of the sensitivity-analysis parameters π_x, α_x and δ_{xt} in the approach proposed by Rosenbaum and Rubin we will show how to obtain the quantities needed for the application of the formula for d_{a_+} in Theorem 1, namely, (i) $\{E(Y|a_1, x, u) - E(Y|a_1, x, u')\}$ and $\{E(Y|a_0, x, u) - E(Y|a_0, x, u')\}$ and (ii) $\{P(u|a_1, x) - P(u|x)\}$ and $\{P(u|a_0, x) - P(u|x)\}$. Given α_x and π_x, we can use the equations

$$\begin{aligned} \alpha_x &= \log\left(\frac{P(a_0|x,u)}{1-P(a_0|x,u)} / \frac{P(a_0|x,u')}{1-P(a_0|x,u')}\right) \\ P(a_0|x) &= P(a_0|x,u)P(u|x) + P(a_0|x,u')P(u'|x) \end{aligned}$$

to solve for $P(a_0|x, u)$ and $P(a_0|x, u')$; note that $P(u|x) = \pi_x$ and $P(u'|x) = 1 - \pi_x$. It then follows from Bayes' rule that

$$P(u|a_0, x) = \frac{P(a_0|x,u)\pi_x}{P(a_0|x)}.$$

A similar procedure can be used to obtain $P(u|a_1, x)$ and thus also $P(u|x)$. Furthermore to obtain $\{E(Y|a_1, x, u) - E(Y|a_1, x, u')\}$ and $\{E(Y|a_0, x, u) - E(Y|a_0, x, u')\}$, we can use the equations

$$\begin{aligned} \delta_{x0} &= \log\left(\frac{1-E(Y|a_0,x,u)}{E(Y|a_0,x,u)} / \frac{1-E(Y|a_0,x,u')}{E(Y|a_0,x,u')}\right) \\ E(Y,U|a_0,x) &= E(Y|a_0,x,u)P(u|a_0,x) + E(Y|a_0,x,u')P(u'|a_0,x) \end{aligned}$$

to solve for $E(Y|a_0, x, u)$ and $E(Y|a_0, x, u')$; note that $P(u|a_0, x)$ and $P(u'|a_0, x)$ have already been obtained. A similar procedure can be used to obtain $E(Y|a_1, x, u)$ and $E(Y|a_1, x, u')$. We can then proceed by using the bias formula for d_{a_+}.

Although the bias formula for d_{a_+} in Theorem 1 can be used to replicate the odds-ratio sensitivity analysis of Rosenbaum and Rubin,[6] the formula in Theorem 1 is considerably more general since, as we have seen in the applications above, it can be applied to binary or continuous outcomes and to binary, categorical or continuous confounding variables and treatment variables. The result of Rosenbaum and Rubin was restricted to binary outcomes.

Relation to Sensitivity Analysis of Lin et al. (1998)

Lin et al.[11] considered settings including binary and continuous outcomes Y, binary treatment A, and binary and continuous unmeasured confounding variable U. They compared the two regression models

$$E(Y|a,u,x) = g(\alpha + \beta a + \gamma u + \theta' x)$$

and

$$E(Y|a,x) = g(\alpha^* + \beta^* a + \theta^{*\prime} x).$$

for linear, log-linear and logistic links g and derived algebraic formulas to relate β and β^* under two possible alternative assumptions. Their first assumption was that U and X were conditionally independent given A. Their second possible assumption was that the mean of U conditional on A and X was additive in A and X i.e. $\mu_{a,x} := E(U|A=a, X=x) = \mu_a + q(x)$. Hernán and Robins[45] showed that the first assumption concerning the independence of U and X conditional on A could not be satisfied if both U and X were causes of A, and thus that the results of Lin et al.[11] concerning the conditional independence assumption could not be employed in those contexts in which the formulas would be most useful, i.e. when both U and X contained confounding variables. VanderWeele[20] showed that the second assumption of Lin et al. concerning additivity held for an entire family of distributions even if both U and X were causes of A. Under this second assumption of additivity, Lin et al.[11] showed that when the conditional distribution of U given A and X is normal with mean $\mu_{a,x} = \mu_a + q(x)$ then the regression coefficients β and β^* were related by

$$\beta = \beta^* - \gamma(\mu_1 - \mu_0)$$

for linear and log-linear links and that this relationship held approximately for a logistic link.

Lin et al.[11] also noted that this relationship would hold for linear link (but not log-linear or logistic) if U were binary (rather than normally distributed) with a conditional mean $\mu_{a,x} = \mu_a + q(x)$.

The results for a linear link follow immediately from the bias formula for $d_{a_+}(x)$ by replacing expressions of the form $E(Y|a,u,x)$ with the linear combination of regression coefficients and by using $E(U|a,x) = \mu_a + q(x)$. See also Cox and McCullagh.[32] Furthermore, from the bias formula for d_{a_+} it also follows that to obtain $\beta = \beta^* - \gamma(\mu_1 - \mu_0)$, the unmeasured confounding variable U does not need to be binary nor to be conditionally normally distribution; all that is needed to obtain $\beta = \beta^* - \gamma(\mu_1 - \mu_0)$, or more generally to obtain $\beta = \beta^* - \gamma(\mu_{a_1} - \mu_{a_0})$, is that the conditional mean of U given X and A is given by $\mu_{a,x} = \mu_a + q(x)$. Not only do all of these results for linear regression follow immediately from Theorem 1, but the bias formula for $d_{a_+}(x)$ can in fact be used to relax the additivity assumption. Without making an assumption about additivity, it follows immediately from the bias formula for d_{a_+} that β and β^* are related by:

$$\beta = \beta^* - \gamma \int_x \int_u \{dP(u|a_1,x) - dP(u|a_0,x)\} dP(x).$$

The bias formulas in Theorem 1 are yet more general than this in that, as we saw in the applications above, unlike the results of Lin et al.[11] and Imbens,[14] Theorem 1 does not presuppose a regression context; furthermore it does not assume that there are no interactions between A, U and X, and does not presuppose any particular functional form.

Relation to Prior External Adjustment Results

As noted above, the bias formulas in Theorem 1 are a generalization of prior bias formulas in the external adjustment literature.[1,8,21] This prior external adjustment literature generally considered the setting of a dichotomous treatment, a dichotomous outcome and a categorical unmeasured covariate. If in Theorem 1, $X = \varnothing$, Y is dichotomous and U is categorical, the formulas given above reduce to the results for the risk difference in Kitagawa[1] and Arah et al.[21] The bias formula results given here thus generalize these prior external adjustment results in two ways. First our results apply not only to dichotomous outcomes but also to continuous outcomes. Second, our results allow for control of some set of measured covariates X. Whereas prior risk difference results compared the average outcome difference unadjusted for X, $E(Y|a_1) - E(Y|a_0)$, to the causal effect $E(Y_{a_1}) - E(Y_{a_0})$, our results compare the average outcome difference adjusted for X, $\sum_x \{E(Y|a_1, x) - E(Y|a_0, x)\}P(x)$, to the causal effect $E(Y_{a_1}) - E(Y_{a_0})$. Most of the prior external adjustment literature effectively presupposed that the analysis was within strata of, or conditional on, X (or that there were no measured covariates to control for). In order to combine results over strata of X, Lee and Wang[46] and Flanders and Khoury[9] for instance propose an assumption of homogeneity of effects across of X; Greenland[15,16] and Arah et al.[21] discuss Bayesian and Monte Carlo method approaches; using risk ratios, Flanders and Khoury[9] also derive an external adjustment formula in cases with both a measured and an unmeasured covariate.