WEB-BASED SUPPLEMENTARY MATERIAL FOR *SENSITIVITY ANALYSIS FOR MATCHED PAIR ANALYSIS OF BINARY DATA: FROM WORST CASE TO AVERAGE CASE ANALYSIS* BY R. HASEGAWA AND D. SMALL

R. HASEGAWA* AND D. SMALL†

Web Appendix A

Time-varying propensity for cellphone use. After conditioning on the time-invariant driver characteristics $\mathbf{X}$, it may be natural to model U_{si} as a time-varying propensity quantile for using a cellphone in the hazard window ($i = 1$) and in the control window ($i = 2$). U_{si} could summarize an arbitrary number of confounding variables that vary between control and hazard windows for driver s. As a quantile, we can think of U_{si} as coming from a uniform distribution on $[0, 1]$. U_{s1} and U_{s2} can conceivably be considered independent since by design a case-crossover study controls for all individual, time-invariant confounders. Now, suppose that we conduct a sensitivity analysis that returns Γ_{sens}. If we interpret this as an average case hidden bias we may want to ask how large we would expect the corresponding worst case hidden bias to be. If we assume the the U_{si} are propensity quantiles that are iid uniformly distributed we can compute a lower bound for the expected worst case hidden bias corresponding to the average case calibrated Γ_{sens}. Let Σ be the set of all permutations of $\{11, 12, \ldots, S1, S2\}$ and let $\sigma \in \Sigma$ be an element in the set. Now define $\gamma^*(\mathbf{U})$ to be the solution to

$$\Gamma_{sens}/(1+\Gamma_{sens}) = \sup_{\sigma\in\Sigma}\left\{\frac{1}{S}\sum_{s=1}^{S}\frac{\exp(\gamma(U_{\sigma(s1)}-U_{\sigma(s2)}))}{1+\exp(\gamma(U_{\sigma(s1)}-U_{\sigma(s2)}))}\right\}. \tag{1}$$

The right hand side of this equation inside the supremum operator is an expression for $\bar{\mathbf{p}}$ under the sensitivity model defined in Section 2.2. The following proposition and corollary show that $\Gamma^* = \exp(\gamma^*(\mathbf{U}))$ is a lower bound for the worst case calibrated hidden bias corresponding to the average case calibrated Γ_{sens} given $\mathbf{U}$.

Proposition 1. *With probability one $\gamma^*(\mathbf{U})$ is the unique solution of (1) and the smallest γ that satisfies*

$$\Gamma_{sens}/(1+\Gamma_{sens}) = \frac{1}{S}\sum_{s=1}^{S}\frac{\exp(\gamma(U_{\sigma(s1)}-U_{\sigma(s2)}))}{1+\exp(\gamma(U_{\sigma(s1)}-U_{\sigma(s2)}))}$$

for some $\sigma \in \Sigma$.

Proof. It suffices to show that the right hand side of (1) is strictly increasing in γ with probability one. Consider $0 \le \gamma_1 < \gamma_2$ and let σ_1 be the permutation that maximizes

$$f(\sigma,\gamma_1,\mathbf{U}) = \frac{1}{S}\sum_{s=1}^{S}\frac{\exp(\gamma_1(U_{\sigma(s1)}-U_{\sigma(s2)}))}{1+\exp(\gamma_1(U_{\sigma(s1)}-U_{\sigma(s2)}))}.$$

Assuming that the U_{si} are iid uniform, $\mathbf{U}$ is nonconstant with probability one. If $\mathbf{U}$ is nonconstant then $f(\sigma_1,\gamma_1,\mathbf{U}) < f(\sigma_1,\gamma_2,\mathbf{U})$ since $U_{\sigma(s1)} - U_{\sigma(s2)} > 0$ for at least one $s = 1,\dots,S$. If we let σ_2 be the permutation that maximizes $f(\sigma,\gamma_2,\mathbf{U})$ then we have that $f(\sigma_1,\gamma_1,\mathbf{U}) < f(\sigma_2,\gamma_2,\mathbf{U})$, completing the proof. □

Corollary 2. *Under the sensitivity model defined in Section 2.2, $\Gamma^* = \exp(\gamma^*(\mathbf{U}))$ is a lower bound for the worst case calibrated hidden bias corresponding to the average case calibrated Γ_{sens} given $\mathbf{U}$.*

Using Corollary 2 we can determine the expected lower bound on the worst case hidden bias corresponding to the average case calibrated Γ_{sens} by computing $\mathbb{E}[\Gamma^*]$ via Monte Carlo estimation. The expectation here is taken over $\mathbf{U} \sim \mathcal{U}[0,1]^{2S}$. Under this propensity quantile model for the unobserved confounders, this procedure can give us a sense of how much less conservative the average case calibration is than the worst case calibration. In the table below we give Monte Carlo estimates and standard errors for $\mathbb{E}[\Gamma^*]$ corresponding to the average case calibrated Γ_{sens} for each of the four control windows.

For each control window we find that the average case interpretation is significantly less conservative than the worst case interpretation. For example, using the worst case calibrated sensitivity analysis we can conclude that if no case-crossover pair was subject to hidden bias larger than 4.15, there is still significant evidence at level

Control Window	Γ_{sens}	$\mathbb{E}[\Gamma^*]$
previous weekday/weekend	4.92	24.76 (0.08)
one week prior	5.53	31.42 (0.11)
previous driving day	4.15	17.66 (0.06)
most active telephone day	2.40	5.81 (0.01)

WEB TABLE 1. Sensitivity analysis for (marginal) $\alpha = 0.05$ and the expected lower bound on corresponding worst case calibrated bias $\mathbb{E}[\Gamma^*]$. Standard errors for Monte Carlo estimates are in parentheses.

$\alpha = 0.05$ that talking on the phone increases the risk of getting in a car accident. In contrast, using the average case calibration we can say there is significant evidence of a treatment effect even if we *expect* that the worst case bias in any case-crossover pair to be greater than 17.66. Notice that for larger values of Γ_{sens} the benefit from using the average case calibration increases. This exercise is not necessarily meant to be a general purpose procedure but rather a numerical illustration of the gain in power that comes from using the average case calibration even under relatively innocuous assumptions about $\mathbf{U}$.

Web Appendix B

Simultaneous sensitivity analysis. The one-parameter sensitivity model introduced in Section 2.2 is often referred to as the *primary sensitivity analysis*. In this model, the association between U_{si} and Z_{si} is controlled by Γ and the stochastic ordering in Equation (5) of the main paper were derived by Rosenbaum (1987) assuming that U_{si} and r_{Csi} have a near perfect relationship but this is not always a plausible assumption – for example, if U_{si} is continuous propensity score for treatment and the outcome is binary. A more general two-parameter sensitivity model was first introduced by Gastwirth et al. (1998) where Δ controls the association between U_{si} and r_{Csi} and Λ controls the association between U_{si} and Z_{si}. For example, if $U_{s1} = 1$ and $U_{s2} = 0$ then the first unit of the pair is Λ times more likely to receive treatment and Δ times more likely to have the positive outcome. This model is known as the *simultaneous sensitivity analysis* and is particularly useful when it is not plausible that U_{si} and r_{Csi} are perfectly correlated. When the outcome is binary – and more generally if y_s, the difference in outcomes in pair s, come from a distribution belonging to Wolfe's semiparametric family (see Wolfe (1974)) – Rosenbaum and Silber (2009) show that Γ can be *amplified* to the two-parameter model (Λ, Δ) by the identity $\Gamma = (\Lambda\Delta + 1)/(\Lambda + \Delta)$ which we refer to as the *amplification curve*. The simultaneous sensitivity model acts as an interpretive aid to the standard one-parameter procedure; We may consider how u_{si} affects the odds of treatment and the odds of positive outcome separately and then use the amplification curve to determine the corresponding Γ with which we can perform a standard one-parameter sensitivity analysis. Like the one-parameter sensitivity model, the simultaneous sensitivity model does not require that the investigator specifies the distribution of U_{si}, only that $U_{si} \in [0, 1]$ for $s = 1, 2, \ldots, S$ and $i = 1, 2$.

When U is not bounded. Theorem 1 in the main paper is free from any modeling decision of the underlying causal mechanism. In that sense, it is very general and allows us to relax the restriction that U lie in the unit interval. This provides flexibility in modeling the unobserved confounders but for Γ, or Δ and Λ in the amplified setting, to retain meaning we will have to standardize the distribution of U in some fashion. For example, we may scale U such that the post matching variance of $U_{s1} - U_{s2}$ is equal to 1. Now suppose that $U_{s1} - U_{s2} = \pm 1$. Then the odds that the treated unit has positive outcome in pair s is

$$\Gamma = \frac{1 + \Lambda\Delta}{(1 + \Lambda)(1 + \Delta)}, \tag{2}$$

which is equivalent to the worst case bias when U was taken to lie on the unit interval. We will refer to this as the one standard deviation (1SD) worst case bias. Related work by Wang and Krieger (2006) considers arbitrary distributions of $W_s = U_{s1} - U_{s2}$ with mean 0 and variance 1 after matching. They show that for any such distribution of $\mathbf{W} = (W_1, W_2, \ldots, W_S)^T$, the population mean of p_s, $\mathbb{E}[p_s]$, is maximized when W_s takes values ± 1 with equal probability. The implication of this result is that when U is scaled appropriately the 1SD worst case bias is asymptotically more conservative than Γ' even when U is not restricted to the unit interval.

References

Gastwirth, J. L., Krieger, A. M., and Rosenbaum, P. R. (1998). Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika* **85,** pp. 907–920.

Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika* **74,** 13–26.

Rosenbaum, P. R. and Silber, J. H. (2009). Amplification of sensitivity analysis in matched observational studies. *Journal of the American Statistical Association* **104,**.

Wang, L. and Krieger, A. M. (2006). Causal conclusions are most sensitive to unobserved binary covariates. *Statistics in medicine* **25,** 2257–2271.

Wolfe, D. A. (1974). A characterization of population weighted-symmetry and related results. *Journal of the American Statistical Association* **69,** 819–822.

*Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340

E-mail address: `raiden@wharton.upenn.edu`

†Statistics Department, The Wharton School, University of Pennsylvania, 400 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340

E-mail address: `dsmall@wharton.upenn.edu`