Supplementary Materials

S1. Matching and Subclassification

Both matching and subclassification can be used to create groups of similar individuals in Step 4 of LSPS.

S1.1 Matching

There are a few criteria that need to be pre-specified for matching. First, the maximum allowed difference between matched individuals, often called caliper. A caliper can be defined at three commonly used scales: propensity score scale, standardized scale, and standardized logit scale. On the standardized scale, the caliper is defined as standard deviations of the propensity score distribution. The standardized logit scale is similar, except that the propensity score is transformed to the logit scale because the PS is more likely to be normally distributed on that scale (Austin, 2011). The default caliper is 0.2 on the standardized logit scale, as recommended by Austin (2011).

Second, the maximum number of persons in the control group to be matched to each person in the treatment group. In its simplest form, 1:1 nearest neighbor matching selects one individual from the control group with the smallest propensity score difference from a given individual in the treatment group. Alternatively, 1:k matching and varied ratio matching can be used. In this study, 1:1 matching is used.

S1.2 Subclassification

Alternative to matching, researchers can use the propensity model to stratify the dataset (Rosenbaum and Rubin, 1984). Subclassification divides individuals into subclasses within which the propensity scores are relatively similar.

Researchers need to determine the number of subclasses and the boundaries of the subclasses. Current convention is to create 5-10 subclasses with equal support (or equal sample size), though more subclasses may be feasible and appropriate if sample size is large. We use 10 subclasses with equal support in this study. Each datapoint is then assigned to a subclass based on its propensity score.

Once the subclasses are defined, LSPS checks that the subclassification achieves covariate balance. Balance is achieved if when we reweigh the data according to the subclasses, each covariate in the treatment and control group follows the same distribution.

To check covariate balance, we first compute the weight for each datapoint w_i to be equal to the reciprocal of the number of datapoints from its treatment group within its subclass. Mathematically, that is,

$$w_i = 1/\Big(\sum_{i}^{N} 1(T_i = t)1(S_i = s)\Big),$$

where the denominator is the number of datapoints that received treatment t in subclass s. Then, the weighted mean of the covariate for treatment group t is

$$\bar{x}_t = \frac{\sum_{i:t_i=t} w_i x_i}{\sum_{i:t_i=t} w_i}, \quad t \in \{0, 1\}.$$

The weighted covariate variance is defined as

$$\sigma_t^2 = \frac{\sum_{i:t_i=t} w_i}{(\sum_{i:t_i=t} w_i)^2 - \sum_{i:t_i=t} w_i^2} \sum_{i:t_i=t} w_i (x_i - \bar{x}_t)^2.$$

With the weighted covariate mean and variance, LSPS computes the standardized mean difference to assess covariate balance.

S2. Outcome Model for Survival Analysis

In the empirical studies of Section 4, we use a Cox proportional hazards model (Cox, 1972) to estimate an unbiased hazard ratio (the outcome Y is time to event). The Cox model is expressed by the hazard function denoted by $h(\tau)$. Within a subclass s, the subclass-specific hazard function $h^s(\tau)$ is estimated as,

$$h^s(\tau) = h_0^s(\tau) \exp(\zeta_s t),$$

where τ is the survival time, $h_0^s(\tau)$ is the baseline hazard at time τ, t is the treatment, and $\exp(\zeta_s)$ is the subclass-specific hazard ratio of the treatment. This expression gives the hazard function at time τ for subjects with treatment t in subclass s.

The parameters in the Cox model are estimated by optimizing the likelihood

$$L(\zeta_s) = \prod_{i:C_i=1 \cap S_i=s} \frac{\exp(\zeta_s t_i)}{\sum_{j:Y_j \geq Y_i} \exp(\zeta_s t_j)},$$

where $C_i = 1$ indicates the occurrence of the outcome.

The hazard ratio can be obtained by reweighting $\exp(\zeta_s)$ by the size of the subclass. However, in practice, due to within-subclass zero counts and finite machine precision, the hazard ratio is estimated by optimizing the Cox partial conditional across subclasses as in the Cyclops R package [69].

S.3 Proof of Theorem 2

We prove Theorem 2, which states that under certain monotonicity assumptions about the effect of the unmeasured confounder on the treatment and on the outcome, the average treatment effect adjusting for the all measured covariates lies between the true effect and the effect adjusting for only the measured confounders.

Let T be a binary treatment, Y be an outcome, U be an ordinal unmeasured confounder, $U \in \{1, \ldots, K\}$, and $\mathbf{X}$ be some measured covariates. Furthermore, let $\mathbf{X}_C$ be the measured confounders, and $\mathbf{X}_C^\complement$ be the complement set of the measured confounders, $\mathbf{X}_C^\complement = \{\mathbf{x} : \mathbf{x} \in \mathbf{X} : \mathbf{x} \notin \mathbf{X}_C\}$. We assume that the measured covariates $\mathbf{X}_C^\complement$ form a noisy measurement of U. We denote this noisy measurement as U'. The relationship between the true unmeasured confounder U and the noisy measurement U', is established by the following misclassification probabilities, $p_{ij} = p(U' = i \,|\, U = j)$, $i, j \in \{1, \ldots, K\}$. Then, $p(U' = i) = \sum_{j=1}^{K} p_{ij} p(U = j)$.

We assume that the misclassification probabilities of U is nondifferential with respect to T, Y, and $\mathbf{X}$; that is, $p(U' = u' \,|\, U = u, Y = y, T = t, \mathbf{X} = \mathbf{x}) = p(U' = u' \,|\, U = u, Y = y', T = t', \mathbf{X} = \mathbf{x}')$ for all $y, y' \in \mathcal{Y}$, $t, t' \in \{0, 1\}$, and $x, x' \in \mathcal{X}$. If $p_{ij} \leq p_{ik}$ and $p_{ji} \leq p_{ki}$ for $j < k < i$, and $p_{il} \geq p_{im}$ and $p_{li} \geq p_{mi}$ for $i < l < m$, then we say that the misclassification probabilities are tapered.

The proof of this theorem relies on Lemma 1-4. We extend Lemma 1-4 in Ogburn and VanderWeele [53] to conditions where identifiability holds by conditioning on measured and unmeasured confounders.

Lemma 1. *If* $\mathbb{E}_{X_C}[Y \,|\, T = 1] \geq \mathbb{E}_{X_C, U'}[Y \,|\, T = 1] \geq \mathbb{E}[Y(1)]$ *and* $\mathbb{E}_{X_C}[Y \,|\, T = 0] \leq \mathbb{E}_{X_C, U'}[Y \,|\, T = 0] \leq \mathbb{E}[Y(0)]$*, or if* $\mathbb{E}_{X_C}[Y \,|\, T = 1] \leq \mathbb{E}_{X_C, U'}[Y \,|\, T = 1] \leq \mathbb{E}[Y(1)]$ *and* $\mathbb{E}_{X_C}[Y \,|\, T = 0] \geq \mathbb{E}_{X_C, U'}[Y \,|\, T = 0] \geq \mathbb{E}[Y(0)]$*, then the treatment effect adjusting for all measured covariates falls between the true effect and the effect adjusting for the measured confounders alone.*

Lemma 2. *If $\mathbb{E}[Y \mid T, X, U]$ and $\mathbb{E}[T \mid X, U]$ are either both nonincreasing or both nondecreasing in U, then $\mathbb{E}_{X_C}[Y \mid T = 1] \geq \mathbb{E}[Y(1)]$ and $\mathbb{E}_{X_C}[Y \mid T = 0] \leq \mathbb{E}[Y(0)]$. If one of $\mathbb{E}[Y \mid T, X, U]$ and $\mathbb{E}[T \mid X, U]$ is nonincreasing and the other nondecreasing in U, then $\mathbb{E}_{X_C}[Y \mid T = 1] \leq \mathbb{E}[Y(1)]$ and $\mathbb{E}_{X_C}[Y \mid T = 0] \geq \mathbb{E}[Y(0)]$.*

Lemma 3. *Suppose that U is nondifferentially misclassified with respect to T and Y. If $\mathbb{E}[Y \mid T, X, U]$ and $\mathbb{E}[T \mid X, U]$ are both nondecreasing or both nonincreasing in U, then $\mathbb{E}_{X_C, U'}[Y \mid T = 1] \geq \mathbb{E}[Y(1)]$ and $\mathbb{E}_{X_C, U'}[Y \mid T = 0] \geq \mathbb{E}[Y(0)]$. If one of $\mathbb{E}_{X_C, U'}[Y \mid T, X, U]$ and $\mathbb{E}[T \mid X, U]$ is nondecreasing and the other nonincreasing in U, then $\mathbb{E}_{X_C, U'}[Y \mid T = 1] \leq \mathbb{E}[Y(1)]$ and $\mathbb{E}_{X_C, U'}[Y \mid T = 0] \leq \mathbb{E}[Y(0)]$.*

Lemma 4. *Suppose that U is nondifferentially misclassified with respect to T and Y with tapered misclassification probabilities. If $\mathbb{E}[Y \mid T, X, U]$ and $\mathbb{E}[T \mid X, U]$ are both nondecreasing or both nonincreasing in U, then $\mathbb{E}_{X_C, U'}[Y \mid T = 1] \leq \mathbb{E}_{X_C}[Y \mid T = 1]$ and $\mathbb{E}_{X_C, U'}[Y \mid T = 0] \leq \mathbb{E}_{X_C}[Y \mid T = 0]$. If one of $\mathbb{E}[Y \mid T, X, U]$ and $\mathbb{E}[T \mid X, U]$ is nondecreasing and the other nonincreasing in U, then $\mathbb{E}_{X_C, U'}[Y \mid T = 1] \geq \mathbb{E}_{X_C}[Y \mid T = 1]$ and $\mathbb{E}_{X_C, U'}[Y \mid T = 0] \geq \mathbb{E}_{X_C}[Y \mid T = 0]$.*

Lemma 2 establishes the relationship between the true expectations and expectations conditioning on the measured confounders. Lemma 3 establishes the relationship between expectations conditioning on all measured covariates and the true expectations. Lemma 4 establishes the relationship between expectations conditioning on all measured covariates and expectations conditioning on the measured confounders. Lemma 2-4 establish the conditions of Lemma 1, and Theorem 2 then follows.

References

P. C. Austin, An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies, Multivariate Behavioral Research 46(3) (2011) 399–424.

D. R. Cox, Regression Models and Life-Tables, JSTOR 34 (2) (1972) 187–220.

E. L. Ogburn, T. J. Vanderweele. Bias attenuation results for nondifferentially mismeasured ordinal and coarsened confounders, Biometrika 100(1) (2013) 241–248.

P. R. Rosenbaum, D. B. Rubin, Reducing Bias in Observational Studies Using Subclassification on the Propensity Score, Journal of the American Statistical Association 79 (387) (1984) 516.

M. A. Suchard, S. E. Simpson, I. Zorych, P. Ryan, D. Madigan, Massive parallelization of serial inference algorithms for complex generalized linear models, ACM Transactions on Modeling and Computer Simulation 23 (2013) 1-17.