UNLEARN

Poster #302409: PROCOVA using Disease Progression Model Predictions vs Propensity Score Matched Synthetic Control Arms

Dave P. Miller

Background

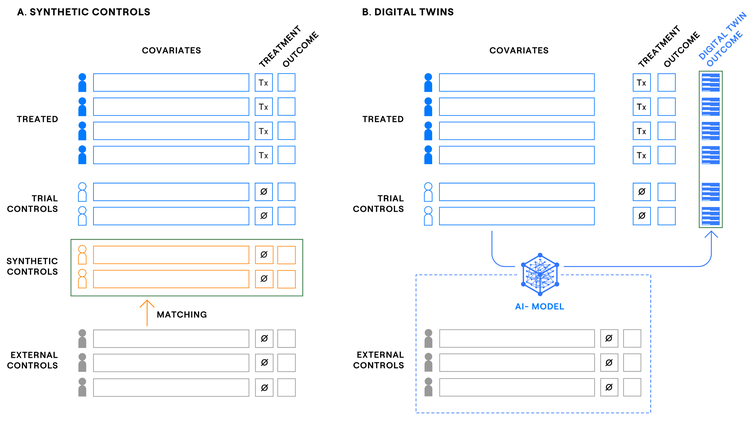
There is a growing need to improve efficiency in clinical trials, lowering costs and making effective drugs available to patients sooner. Leveraging external datasets of untreated patients to complement trial data, improving power and efficiency, is becoming increasingly attractive, especially with the access to large electronic datasets.

One approach is the use of a synthetic control arm (SCA). In this approach, a single-arm clinical trial with all patients receiving the active treatment is supplemented by a SCA taken from an external data source. Propensity score matching (PSM) is a common technique advocated to control for differences between the treatment arm and the SCA (a subset of the external data source).

An alternative and novel approach is using Digital Twins (DTs) and Prognostic Covariate Adjustment (PROCOVA). In this approach, the external data source is used to build a disease progression model (DPM). Patients are still randomized into a treatment and a control arm, just as a typically randomized clinical trial (RCT) does. The added value comes from the use of the DPM to generate DTs for each patient in the treatment or control arm. The Digital Twins are clinical predictions, based on patients' baseline data alone, for what would have happened to each patient if randomized to the placebo arm. When the Digital Twin predictions are entered into an ANCOVA model as a covariate (i.e., applying PROCOVA), the power to find a treatment effect increases because of the variance explained by the Digital Twin.

While the adjectives "synthetic" and "digital" may sound similar, Synthetic Controls add patients to the trial, while Digital Twins add data about the existing patients in the trial. In fact, with synthetic controls, it's the arm of the trial that's synthetically created, but the external data that make up the control arm are real patients. Thus, if one had 200 patients in a single-arm clinical trial and added one SCA with 100 synthetic controls, the study dataset to be analyzed now has 300 real patients. On the other hand, Digital Twins are computer-generated from DPM, so they are not real patients. The Digital Twin data provides additional predictive information on each patient in the clinical trial. From a dataset perspective, Digital Twins add columns to the trial dataset and Synthetic Controls add rows, as shown by the green boxes in Figure 1.

Figure 1: Synthetic Controls vs Digital Twins



Simulation

We randomly generated a cohort of 200 clinical trial patients and a cohort of 4000 external dataset patients. Both cohorts had a random measured covariate, a random unmeasured covariate, and an outcome variable. The outcome variable was generated as a function of the measured covariate, the unmeasured covariate, random error, and, for treated clinical trial patients, a treatment effect.

To simulate the data, two other parameters were varied; the correlation between the measured and unmeasured covariate (Rmu), and the relative contribution of the unmeasured covariate versus the measured covariate (Wu) in predicting which patients were in the clinical trial cohort versus the external dataset . A value of 1 indicated an equal contribution. A value of 0 indicated that group membership was only predicted by the measured covariate.

Once the clinical trial data and external data were simulated, results for both propensity-matched synthetic controls and Digital Twins with PROCOVA were generated. For both the propensity matching and the generation of Digital Twins, the unmeasured covariate was invisible (i.e., neither the propensity matches nor the DPM used to generate Digital Twins could not utilize the unmeasured covariate).

For each simulation, a final dataset with 300 patients (200 treated patients and 100 propensity-matched synthetic controls, i.e., the SCA) were used to compute a treatment effect and a p-value. Similarly, for each simulation a final dataset with 200 patients (100 treated patients with a Digital Twin generated for each patient and 100 patients receiving placebo with a Digital Twin generated for each patient) were used to compute a treatment effect and a p-value using PROCOVA. To assess type I error, the treatment effect was set to zero, and the percentage of simulations with a P value greater than 0.05 was computed.

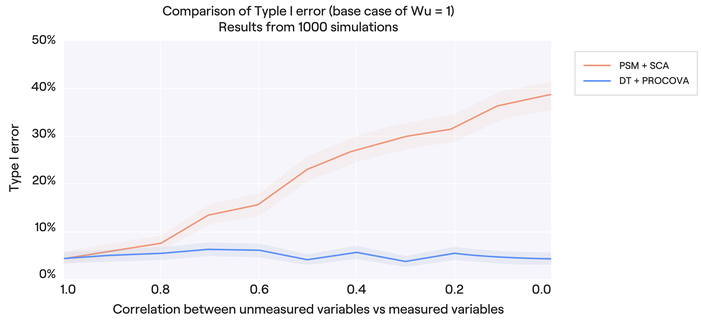
For the base case, the correlation between the measured and unmeasured covariate (R) was set to 0.4 and the relative contribution of the unmeasured and measured covariate (Wu) was set to 1.

Results

For the base case of Rmu=0.4 and Wu=1, over 1000 simulations, the type I error for propensity score matching (PSM)+SCA was 26% ± 2% and the type I error for Digital Twins (DT) +PROCOVA was 5% ± 1%.

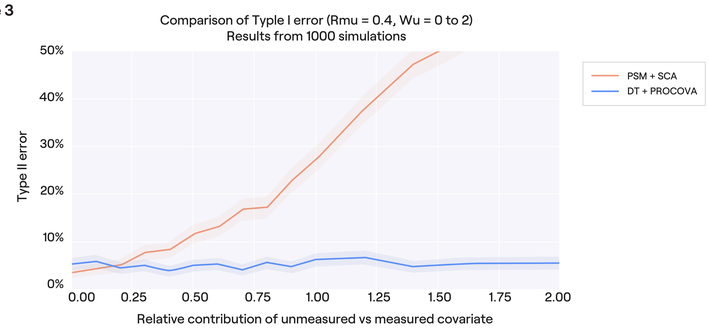
In Figure 2, allowing the correlation between the measured and unmeasured covariate (Rmu) to vary, we see that type I error control is retained when Rmu=1, but type I error control begins to be lost even at a very high correlation of 0.8.

Figure 2



In Figure 3, we set Rmu=0.4 and allow the relative strength of the unmeasured covariate relative to the measured covariate to vary. The type I error for DT+PROCOVA remains constant close to 5%, whereas the type I error for PSM+SCA is 5% or below only when Wu=0. Even when the relative strength of the unmeasured covariate is only a quarter of that of the measured covariate, the type I error begins to rise above 5% with PSM+SCA. When Wu=1.5, the type I error is 50%, which basically translates into a 50/50 chance of erroneously approving an ineffective drug.

Figure 3



As shown in Figure 4 below, both methods reduce type II error, using our base case (Rmu=0.4, Wu=1), though the PSM+SCA approach reduces type II error to nearly zero.

Figure 4

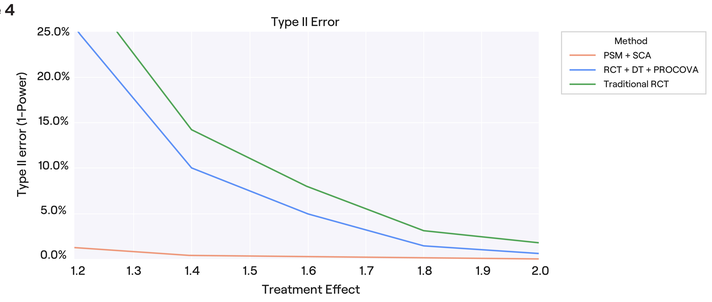
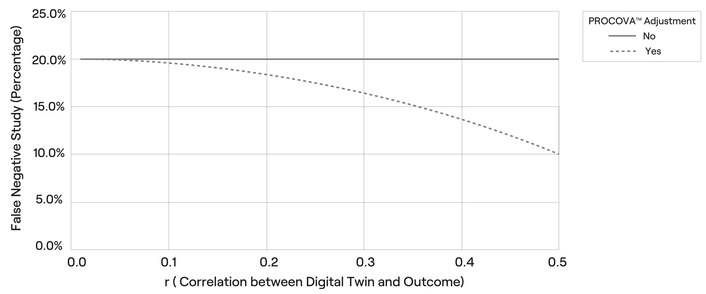


Figure 5



The relationship between DT+PROCOVA and power depends on an additional variable that was not directly included as a variable in our simulations, which is the correlation between the DT and the observed outcome variable. That relationship is shown in Figure 5 below. As the correlation increases from .3 to .4 to .5, the relative reduction in the type II error increases from approximately 20% to 30% to 50%.

Conclusions

Propensity score matching plus synthetic control arms (PSM+SCA) cannot control type I error unless there is either a very high correlation between the measured and unmeasured confounders (e.g., >0.8) or if unmeasured confounders have a minimal effect on group membership versus measured confounders (e.g., <.25). These are very strong assumptions, as correlations over 0.8 are rare and unmeasured confounders often play an important role in distinguishing clinical trial cohorts vs external controls. Most importantly, because unmeasured confounders are unmeasured, the assumptions related to the unmeasured confounders are completely unverifiable.

Digital Twins plus PROCOVA (DT+PROCOVA) provide an opportunity to improve power or reduce sample size without affecting type I error. This approach of adjusting for covariates is consistent with recent FDA guidance. While DT+PROCOVA does not provide as large a power gain as PSM+SCA, it provides a safe power gain that does not rely on unverifiable assumptions. The fact that using PSM+SCA translates into huge gains in power without having a control arm of patients receiving placebo and without randomization is incredibly appealing, but the likelihood of a biased (and incorrect) finding is surprisingly high, even with moderately high correlations between measured and unmeasured confounders.