Matching on E[Y(1)|X]

Suppose you had models for *both* potential outcomes: under treatment and under control. Matching on the expected control outcome (prognostic score) may convey some desirable statistical properties. Should you also match on expected treatment outcomes? Interestingly, maybe not. If the estimand is the sample average treatment effect among the treated, then the control potential outcomes under the treatment may not matter. As an illustration, let $\{Y_i(0), Y_i(1)\}_{i=1}^{n_1}$ represent the potential outcomes of each of the treated individuals in a sample and $\{Y_{i'}(0), Y_{i'}(1)\}_{i=1}^{n_1}$ represent the potential outcomes of each of n_1 matched control individuals. The sample average treatment effect among the treated is:

$$\tau^{SATT} = \sum_{i=1}^{n_t} Y_i(1) - Y_i(0)$$

If we estimate this by

$$\hat{\tau}^{SATT} = \sum_{i=1}^{n_t} Y_i(1) - Y_i'(0)$$

Then our error is:

$$\tau^{SATT} - \hat{\tau}^{SATT} = \sum_{i=1}^{n_t} Y_i(0) - Y_i'(0)$$

This doesn't actually depend on how close $Y_{i'}(1)$ is to $Y_i(1)$. Inessence, since we only care about the control individuals as proxies for understanding the control potential outcomes of the treated individuals, the unobserved control potential outcomes, $\{Y_i'(1)\}_{i=1}^{n_t}$ were inconsequential. Of course, this might change if you were more interested in a more symmetric estimand, such as the ATE.