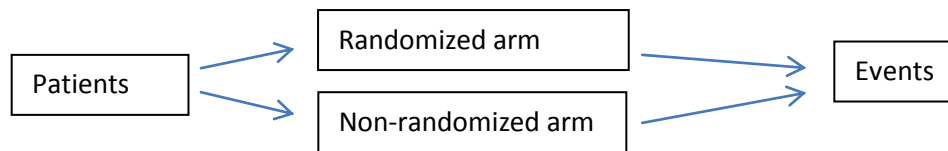


Variables:

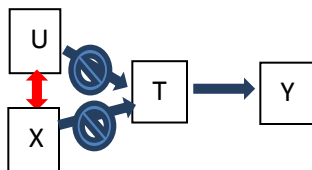
T: 1 if treated, 0 otherwise
Y: Disease outcome
X: Measured confounders such as patients' own clinical risk factors
U: Unmeasured confounders such as patients' unmeasured risk factors that are related to disease

Design:



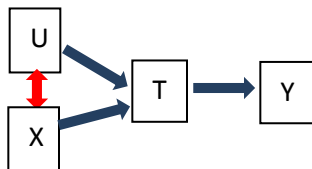
#Randomized arm:

If T is randomized, then T is not confounded with X and U. And hence T has the unbiased effect on Y. Simple regression or correlation analysis will be sufficient to assess the effect of T on Y. In this case, probability of receiving treatment is unconditional on any covariates, that is, $P(T=1 | X,U) = P(T=1)$, which is usually 0.5 (tossing coin, computer randomization).



#Non-randomized arm:

However, if T is confounded by X and U, T will not have unbiased effect on Y. In other words, $P(T=1 | X,U)$ needs to be counted for the analysis.



Analysis strategy:

Since the non-randomized arm is subject to the bias due to unmeasured confounders (residual confoundings), it is critical to measure as many confounders as possible. Once all possible confounders are measured, these variables will be summarized in propensity score, which is the patient-level propensity of receiving treatment given all measured confounders: $P(T=1 | X)$. In case of failure to measure critical confounders, we can conduct sensitivity analysis for the residual confounding.

Propensity score methodologies have been used to address such non-randomized feature of clinical studies. For non-randomized arm, we can do matching, weighting and stratification based on the propensity scores to remove the bias due to self-selection.