Up-Front Matching for Prospective Observational Studies to Mimic Randomization

Ibrahim Turkoz, Ph.D.[1], William H. Olson, Ph.D. [2]

[1]Statistics & Decision Sciences, Janssen Research & Development, LLC; [2] WHO Statistical Consulting, LLC

ABSTRACT

Background: Prospective observational studies (POS's) of the comparative effectiveness of drugs are often based on data from databases created for clinical practice and not research purposes. Federal government/regulatory entities are examining the use of real-world data and evidence for regulatory purposes. To date their emphasis is on ensuring the quality of real-world data and evidence.

Methods: Using the propensity score based on selected covariates in historical claims data, up-front matching (1) enrolls only patients whose propensity score is in the common support; and (2) uses frequency matching based on propensity score strata. Up-front matching is compared to random sampling in a simulation.

Results: Across 500 simulated studies, each with 200 patients per group, the distribution of the standardized mean difference (SMD_{200}) for each of the selected covariates with up-front matching is well approximated by the normal distribution one would see with randomization (N(0, 2/200)), and with random sampling it is well approximated by *N(std.diff,* 2/200), where *std.diff* is the standardized difference in means of the populations from which random samples are taken.

Conclusions: For selected covariates, up-front matching assigns patients to treatment as if they are being randomized to treatment, whereas with no up-front matching the treatment assignment mechanism is simply what one would see with random sampling.

HIGHLIGHTS

- Matching at enrollment for POS's has been used extensively for two situations: (1) direct patient level matching when there is a very small number of categories of patients; and (2) when there is more than just a small number of categories of patients, propensity score matching has been used, but this requires having all potentially enrolled patients and their covariate values in hand before enrollment starts.
- Up-front matching has been developed to handle situations that meet neither of these conditions: there is more than a small number of categories of patients and all potentially enrolled patients with known values of these categories are not available at the outset of the study. Patients are to be recruited in an ongoing basis as they present to investigators, a very common method of enrollment.

OBJECTIVE

For POS's that are designed to compare 2 drugs or classes of drugs, develop a recruitment method, called "up-front matching", that will result in better balance for selected baseline covariates in the patient populations enrolled.

METHODS

Up-Front Matching Methodology

For concreteness up-front matching will be described for determining the average treatment effect in the treated (Drug A) compared to Drug B -similar considerations apply to other causal estimands. Up-front matching is part of the design of the study and does not depend on the outcomes – this follows the recommendations of Imbens and Rubin[1] and Rosenbaum[2]. The key is to use inexpensive covariates to guide enrollment. Let S denote a set of inexpensive covariates in the claims database and n(s) denote the propensity to assign a patient with covariate values **s** to Drug A based on the claims database. The goal of up-front matching is to create enrolled populations in the POS with (1) a higher percentage of patients in the common support; and (2) balance across the inexpensive covariates in the final enrolled populations. The propensity score is used here for its balancing properties[3]. Up-front matching is performed as follows:

a-) Determine the common support based on the propensity score (Π(s)) distributions of the two treatment groups in the claims database.[4]

b-) Determine the deciles of the propensity score distribution of Drug A; we are interested in enrolling patients in both groups who have pretreatment characteristics like those who were treated with Drug A.

c-) As patients are considered for enrollment, enroll only those whose propensity score is in the common support.

d-) Specify a quota of patients to be enrolled for each treatment group in each decile. For our illustration we take this quota to be the same for each treatment-decile combination – this is stratified matching (on the propensity score), a form of statistical sampling that is alternatively known as frequency matching.

e-) For each treatment group once the quota in a treatment-decile group has been filled that treatment-decile is closed to further enrollment.

f-) For each treatment group continue enrolling patients until all treatment-decile quotas for that treatment group have been met.

Illustrative Simulation

Consider an RWE study of the comparative effectiveness of injectable atypical anti-psychotics ("Drug A" - IAP) compared to oral atypical anti-psychotics ("Drug B" - OAP). Patients and the values of their baseline covariates were selected from the Truven Marketscan Medicaid database from January 1, 2011 through December 31, 2016. The sample size for each treatment group in the simulated POS is 200.

A preliminary propensity score model was determined using the R package *twang.*[5,6] that uses boosted logistic regression. Two recruitment schemes were simulated: (1) up-front matching following steps a) through f) given above; and (2) a comparison recruitment scheme which enrolled all patients regardless of their propensity score until the specified sample sizes were met. For each simulated study, the empirical standardized mean difference, SMD, was captured.

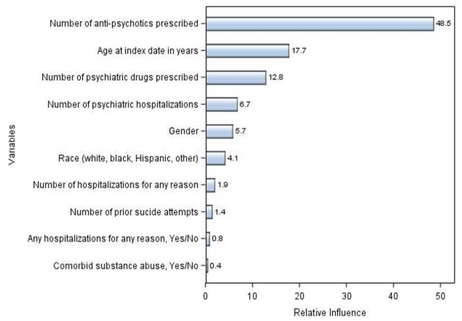
$SMD_n = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{1}{2}(s_1^2 + s_2^2)}}$. In our simulations *n* equals 200.

The POS was simulated 500 times for each matching method (up-front matching and no up-front matching) – each time calculating the SMD_{200}'s for all of the selected covariates. The distribution of the SMD_{200} under up-front matching is compared to what one would see if IAP patients were randomly assigned to IAP and OAP. Based on asymptotic approximations, under randomization the distribution of SMD_{200} is asymptotically normal with *mean 0* and *SD of* $0.1 = \sqrt{\frac{2}{200}}$.

RESULTS & DISCUSSION

The baseline characteristics deemed to be of interest and ascertainable in both the claims database and in patients who are to be considered for enrollment at investigative sites are given in Figure 1 along with their relative importance's based on the preliminary propensity score model.

Figure 1. Matching Variables and Their Relative Influence



Note: All non-demographic variables are based on the 1-year follow-back period.

Table 1 summarizes the final covariates for the patient populations (IAP and OAP) that were used in the simulated POS's. The 3 largest SMD_{200}'s in percent are 45.7%, 20.0%, and -19.4% for number of antipsychotics prescribed, number of psychiatric hospitalizations, and gender. The SMD_{200}'s for the remaining covariates are -3.0% for age, 10.8% for number of all psychiatric drugs prescribed, and -13.8% to 14.5% across races.

Table 1 - Summary of Propensity Score Covariates in Claims Database

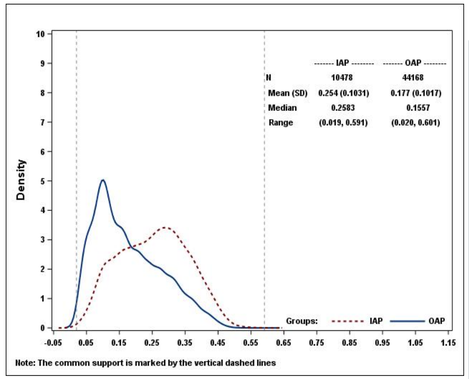
Variables	IAP (n=10478)		OAP (n=44168)		Standardized Difference in Means[1]
	Mean	Variance	Mean	Variance	
Number of anti-psychotics prescribed (*n.antipsychotics*)	1.74	1.69	1.18	1.40	0.457
Age at index date, years (*age*)	37.79	171.19	38.19	193.71	-0.030
Number of psychiatric hospitalizations (*num.psych.hosp*)	0.78	1.91	0.53	1.35	0.200
Number of psychiatric drugs prescribed (*n.all.psychiatric.drugs*)	3.97	8.64	3.66	8.57	0.108
Gender (*male*)	0.41	0.24	0.51	0.25	-0.194
Race (*white*)	0.36	0.23	0.43	0.24	-0.138
Race (*black*)	0.48	0.25	0.41	0.24	0.145
Race (*Hispanic*)	0.02	0.02	0.02	0.01	0.003
Race (*other*)	0.15	0.12	0.15	0.13	-0.013

IAP: injectable atypical anti-psychotics; OAP: oral atypical anti-psychotics. Variable names are listed within the parentheses.

(1) The difference in the means (IAP minus OAP) divided by the square root of the average of the variance in the 2 groups.

Figure 2 gives densities of the propensity score distributions for the IAP and OAP groups based on the final propensity score model.

Figure 2. Distribution of Propensity Scores for IAP and OAP Groups



Note: The common support is marked by the vertical dashed lines

Table 2 gives the means and standard deviations of the SMD_{200} by matching method for the final covariates across the 500 simulated POS's for that method.

- The mean SMD_{200} for each of the selected covariates in the up-front matched population was reduced to levels that would be seen with randomization (i.e, approximately 0) while with no up-front matching they remained essentially as seen in the full observational database.
- With no up-front matching the 3 largest mean SMD_{200}'s among baseline covariates were 45.9%, 20.6%, and -19.8% for number of antipsychotics prescribed, number of psychiatric hospitalizations, and gender, while the SMD_{200}'s for these variables in the population from which patients are enrolled (Table 1) are 45.7%, 20.0%, and -19.4%; the mean SMD_{200}'s for these variables under up-front matching are -0.2%, 0.1%, and 0.9%.
- The standard deviations of the SMD_{200}'s under up-front matching and no up-front matching were essentially what one would see if (1) IAP patients were randomized to IAP or OAP on the one hand; and (2) were selected randomly from each treatment group on the other hand.
- For 2 methods of treatment assignment the validity of the asymptotic normal approximations to the distribution of SMD_{200} is supported by (1) the results in Table 1, Table 2, and (2) normal quantile-quantile plots which are not presented here.

Table 2 - Mean and Standard Deviation of SMD_{200}[1] Under No Up-Front Matching and Up-Front Matching

Variables	No Up-Front Matching		Up-Front Matching	
	Mean SMD_{200}	Standard Deviation of SMD_{200}	Mean SMD_{200}	Standard Deviation of SMD_{200}
Number of anti-psychotics prescribed (*n.antipsychotics*)	0.459	0.102	0.002	0.079
Age at index date, years (*age*)	-0.030	0.102	0.012	0.099
Number of psychiatric hospitalizations (*num.psych.hosp*)	0.206	0.097	0.001	0.096
Number of psychiatric drugs prescribed (*n.all.psychiatric.drugs*)	0.109	0.097	-0.009	0.098
Gender (*male*)	-0.198	0.105	0.009	0.093
Race (*white*)	-0.137	0.098	0.002	0.095
Race (*black*)	0.139	0.098	0.000	0.101
Race (*Hispanic*)	0.008	0.099	-0.002	0.105
Race (*other*)	-0.010	0.106	-0.001	0.101

(1) Standardized mean difference based on samples of size 200 across 500 studies.

- Although up-front matching is based on only a subset of covariates, it is anticipated that it will provide a database that enables more robust and efficient estimates of treatment effect than using no matching at enrollment.
- These benefits are desirable even in POS's not intended for regulatory purposes but will be even more valuable for studies whose results become part of the evidence for regulatory decision making – their results will be more credible and there is the potential for significant cost efficiency in generating the data.
- In addition to potential efficiency gains based on balance there is the real possibility that the percentage of patients not in the common support will be relatively substantial, and the cost savings in not following such patients in a POS could be substantial.

CONCLUSIONS

- Our simulation illustrates a major benefit of up-front matching: it creates populations of patients whose balance on the covariates for which matching was implemented is comparable to what would be achieved with randomization.

REFERENCES

1. Imbens GW, Rubin DB. Causal Inference for Statistics, Social, and Biomedical Sciences, An Introduction. New York: Cambridge University Press, 2015.
2. Rosenbaum PR. Design of Observational Studies. New York: Springer Science+Business Media, 2010.
3. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55.
4. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci 2010;25(1):1-21.
5. twang: Toolkit for Weighting and Analysis of Nonequivalent Groups. Available at: https://CRAN.R-project.org/package=twang.
6. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. Stat Med 2010;29(3):337-46.

Acknowledgements: The authors thank Charlotte Carroll (Janssen Real World Market Analytics while this work was performed, now with Value Evidence Outcomes; Data, Methods, and Analytics, GlaxoSmithKline) and Jack Sheehan (Real-World Value and Evidence, Janssen Scientific Affairs, LLC) for their assistance in constructing the database.

Disclosures: Ibrahim Turkoz is an employee of Statistics & Decision Sciences, Janssen Research & Development, LLC, and William Olsen is a former Janssen Scientific Affairs, LLC employee, and is currently an employee of WHO Statistical Consulting, LLC.
Presented at the ISCTM 16th Annual Scientific Meeting, 19-21 February 2020, Grand Hyatt, Washington DC