Propensity weighted approach to control the confounding effect of unbalanced distribution of placebo responders on the treatment effect in psychiatric clinical trials: an artificial intelligence driven trial simulation study

Roberto Gomeni[1], Seth C. Hopkins[2], Françoise Bressolle-Gomeni[1] and Maurizio Fava[3]

[1]Pharmacometrica, La Fouillade (France), [2]Sumitomo Pharma America, Inc., Marlborough, Massachusetts, USA

[3]Department of Psychiatry, Massachusetts General Hospital, and Harvard Medical School, Boston, MA, USA

Each author have no actual or potential conflict of interest in relation to this presentation

BACKGROUND

Randomized, placebo-controlled clinical trials (RCT) are the gold-standard approach for assessing treatment effect (**TE**). The uncontrolled baseline distribution of individuals' **probability of improvements due to positive expectancies (PE)** of treatment outcomes can lead to biased estimates of TE as conventional study designs and statistical approaches don't account for unbalanced baseline distribution of confounding factors such as PE. A novel methodology was proposed for assessing TE conditional to PE based on an artificial intelligence (**AI**) approach. The inverse of PE was used as weight in the MMRM analysis (**Propensity Scores Weighted analysis or PSW**) to assess TE by controlling the potential confounding effects of unbalanced PE distributions [1,2].

The objective of the present analysis was to compare the outcomes of PSW and reference **non-weighted analysis (NW)** using simulated RCTs characterized by different baseline simulated PE distributions to assess potential risk of inflating Type I/II errors due to unbalanced distribution PE, and to identify alternative and more generalizable approaches for analyzing and reporting RCT results.

DATA

The data of two RCTs (study 810 and 874) were used as templates for the Clinical Trial Simulation. The two studies were randomized, double-blind, parallel-group, placebo-controlled studies evaluating efficacy and safety of paroxetine CR (12.5 and 25mg) versus placebo in patients with major depressive disorder [3].

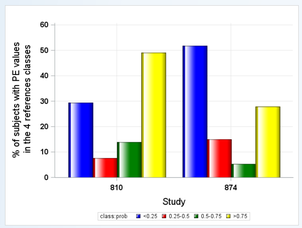

***Fig 1**. Studies 810 and 874: % of subjects with PE values in the 4 reference classes*

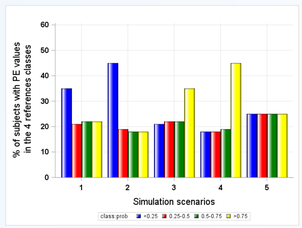

***Fig 2**. % of subjects with PE values in the 4 classes of the 5 simulation scenarios*

CLINICAL TRIAL SIMULATION

The simulation was organized into 2 steps:

1. Estimation of individual PE using AI applied to the individual HAMD-17 item scores estimated at randomization and baseline as potential predictors of PE. **Fig 1** shows the % of subjects with PE probability values in 4 selected classes: **class 1: <0.25%, class 2: 0.25%-0.5%, class 3: 0.5%-0.75%, and class 4: >0.75**% in the studies 810 and 874.
2. Assessment of the impact of different distributions of PE on TE in 5 simulation scenarios (S1 to S5) characterized by different proportions of subjects in the 4 classes of PE. A total of 510 subjects were included in each simulated RCT (**Fig.2**). The subjects were randomly allocated in each simulation scenario according to the following criteria:
   - *S5: 25% of subjects in each class*
   - *S4: 45% of subjects in class 4 (>0.75)*
   - *S3: 35% of subjects in class 4 (>0.75)*
   - *S2: 45% of subjects in class 1 (<0.25)*
   - *S1: 35% of subjects in class 1 (<0.25)*

The same proportion of subjects was allocated in the remaining classes for scenarios 1 to 4 for a total of 170 subjects/arm. The simulations were conducted using a **bootstrap methodology**, based on a Monte-Carlo technique, to constructs synthetic samples by resampling with replacement from the observed experimental samples. For each simulation scenario**, 20 in-silico trials were randomly generated.**

RESULTS

The good predictive performance of the AI models for the PE estimation (Step 1) indicated that:

Study 810: **ROC AUC** (95% CI) = **0.81** (0.64-0.97), Specificity = 0.88, Sensitivity = 0.75, Accuracy = 0.79, and Precision = 0.94;

Study 874: **ROC AUC** (95% CI) = **0.88** (0.77-1.00), Specificity = 0.91, Sensitivity = 0.83, Accuracy = 0.88, and Precision = 0.88.

The plots of the longitudinal HAMD-17 changes from baseline scores observed and by simulation scenario are presented in **Fig 3** for study 810.

Data of each simulated trial were analyzed using the PSW and NW approaches (**Table 1**).

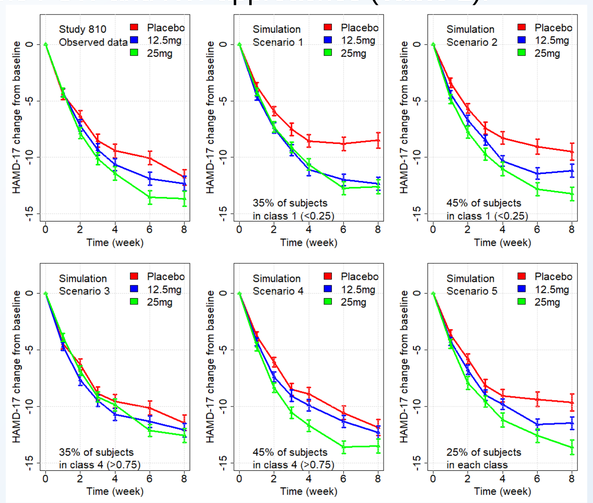

***Fig 3.** Mean (± StdErr) HAMD-17 change from baseline. Top left panel: observed HAMD-17 change from baseline. Remaining panels: HAMD-17 changes from baseline for a study randomly selected among the 20 trials simulated in each scenario*

NW analysis: the TE estimates were statistically different in the different simulation scenarios according to the non-overlapping 95% confidence intervals:

- study 810 TE ranged from 1.3 to 3 for TE1 and from 2.4 to 4.7 for TE2;
- study 874 TE ranged from -0.3 to 1.7 for TE1 and from -0.9 to 1.8 for TE2;
- In the two RCTs larger was the number of subjects in class 4 (PE > 0.75), lower was the estimated TE value, and larger was the number of subjects in class 1 (PE < 0.25), larger was the estimated TE value.

PSW analysis: the TE estimates were very closed and not statistically different in the different simulation scenarios according to the overlapping 95% confidence intervals:

- study 810 TE ranged from 4.1 to 4.7 for TE1 and from 5.1 to 5.9 for TE2;
- study 874 TE ranged from 3.3 to 3.9 for TE1 and from 3.4 to 4.5 for TE2;
- In the two RCTs, the TEs estimated were substantially insensitive to the distribution of PE in class 4 and 1.

	NW		PSW	
Study 810 Simulation Scenarios	**TE1 12.5mg**	**TE2 25mg**	**TE1 12.5mg**	**TE2 25mg**
S1	2.9 (2.4,3.3)	4.7 (4.3,5.0)	4.6 (4.1,5.0)	5.9 (5.5,6.3)
S2	3.2 (2.9,3.5)	4.6 (4.2,5.0)	4.7 (4.3,5.0)	5.7 (5.3,6.2)
S3	2.0 (1.5,2.4)	3.1 (2.7,3.6)	4.1 (3.5,4.6)	5.3 (4.8,5.8)
S4	1.3 (1.0,1.6)	2.4 (2.1,2.7)	4.2 (3.7,4.8)	5.1 (4.5,5.7)
S5	2.5 (2.0,3.0)	3.9 (3.6,4.2)	4.1 (3.6,4.6)	5.2 (4.8,5.7)
Mean	**2.4**	**3.7**	**4.3**	**5**
Range	1.3, 3.2	2.4, 4.7	4.1, 4.7	5.1, 5.9
Study 874 Simulation Scenarios	**TE1 12.5mg**	**TE2 25mg**	**TE1 12.5mg**	**TE2 25mg**
S1	0.3 (-0.2,0.8)	1.3 (0.9,1.7)	3.4 (2.9,3.9)	4.4 (4.1,4.6)
S2	1.0 (0.7,1.3)	1.8 (1.5,2.1)	3.9 (3.5,4.2)	4.5 (4.2,4.8)
S3	-1.1 (-1.4,-0.7)	-0.3 (-0.6,0.1)	3.3 (2.9,3.7)	3.4 (3.0,3.9)
S4	-1.7 (-2.1,-1.4)	-0.9 (-1.4,-0.4)	3.3 (2.8,3.7)	3.4 (2.9,3.9)
S5	-0.3 (-0.6,0.0)	0.6 (0.3,1.0)	3.5 (3.1,3.9)	3.7 (3.4,4.1)
Mean	**-0.3**	**0.5**	**3.5**	**3.9**
Range	-1.7, 1	-0.9, 1.8	3.3, 3.9	3.4, 4.5

***Table 1**. Mean (95% confidence interval) of the estimated TEs in the 20 replicates of the RCTs by study, treatment and simulation scenarios*

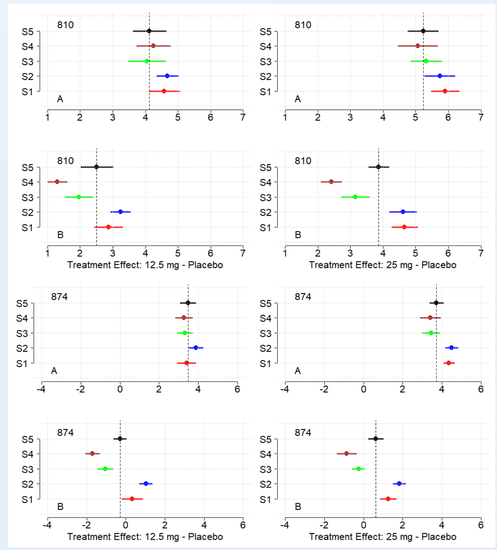

***Fig 4**. Mean TEs (95% confidence intervals) estimated with PSW (panel A) and NW (panel B) methodologies in the 5 simulation scenarios for the CTSs conducted using the 810 and the 874 reference data*

CONCLUSION

Propensity score is a methodology used in observational studies for improving treatment comparison by adjusting data for potentially confounding baseline factors. The simulations indicated that using the NW analysis the risk of inflated Type I error increases with the increase of PE and the risk of Type II error increases with the decrease of PE. **The PSW analysis provided better control of baseline unbalance in PE distribution by providing estimates of TE independent from PE distributions**. These findings emphasize the potential role of PSW as reference approach for analyzing RCTs.

REFERENCES

1. *Psychiatry Res. 2023 Sep;327:115367*
2. *Transl Psychiatry. 2023 Dec 14;13(1):388*
3. *Neuropsychopharmacology. 2015;40:2588-95.*

Corresponding author: roberto.gomeni@pharmacometrica.com