Project: Diagnostics for Informative Censoring (2015-0617)
Authors: Jackson JW, Schnellinger E, Valeri L, Henderson D.

Objective Obtaining intention-to-treat (ITT) effect estimates on symptom reduction and tolerability requires that investigators impute the missing outcomes[1,2]. We applied recently developed data visualizations from the causal inference literature[17] that describe, over time, how dynamic characteristics of patients who dropout differ from those who remain to a placebo-controlled efficacy trial of paliperidone extended release (NCT00397033)[20]. Our objectives were to: (a) describe the dropout mechanism; (b) construct weights that would enable either an ITT comparison of final PANSS scores; (c) use data visualization tools to assess the quality of those weights; (d) report the results before and after applying the weights.

Methods We used visit-specific data to construct a person-time file (target days 0, 4, 8, 15, 22, 29, and 43) which included the assigned treatment arm and data on baseline and time-varying characteristics. Study discontinuation was indexed at the last visit where data was provided; thus, patients providing data at all visits were considered as having completed the study. We excluded the 0-4% those who did not initiate their first study medication and censored patients when they ended double-blind treatment. From a list of potential covariates, we chose those that were independently associated with the outcome or study discontinuation using linear and logistic regression (see Table 1). Fully conditional specification[31,32] was used to impute the relatively minor missing data due to non-response (up to 2% for each covariate in both datasets) using PROC MI in SAS. Generalized additive models were fit on the imputed data to obtain the predicted probability of assigned treatment arm, with and without baseline covariates as predictors, with cubic splines to smooth predictions from continuous covariates. Generalized additive models were also fit on the imputed data to obtain the predicted probability of study discontinuation at visit m. We used these models to construct inverse probability of censoring weights (IPCW) as described in Cole and Hernan 2008.[24]

We plotted the standardized mean difference of each time-varying covariate comparing the censored to uncensored at each visit, both before and after applying the weights to the data. We used GEE to fit unweighted and weighted linear regression models for change in PANSS score from baseline through day 43, both before and after weighting, again assuming an independence covariance matrix. In both applications, a robust-variance sandwich estimator was used to obtain the standard errors of model parameters.[26] P-values from multiple comparisons were adjusted by the Hochberg procedure. All data management and analysis was carried out SAS version 9.3 via the SAS Drug Development Clinical Trial Drug Transparency Platform. All plots were produced using ggplot2 in R. This study was conducted under the approval of the Institutional Review Board at the Harvard T.H. Chan School of Public Health, following all relevant guidelines and protocols.

Results
We obtained 307 patients who were randomized and initiated their assigned treatment. With the exception of race, the distribution of characteristics at baseline was similar (Table 1). There were 99 patients who ended double blind treatment early (of whom 9 provided data at the end of the study but we censored). As expected, dropout was lowest in the high-dose arm (22%), highest in the placebo arm (41%), and was most severe during the first two weeks of follow-up, peaking at target day 15 (Figure 1). The most common reported reason for dropout was lack of efficacy (26% to 50%). These numbers are largely consistent with the lead publication from this study.[20]

The associations between dropout and time-varying symptoms were quite strong (Figure 1). For example, on target day 15 where most of the dropout occurred, many mean differences in PANSS score decline and CGI-Severity were more than a standard deviation apart for the placebo and low-dose arms, and ranged from one to three fifths of a standard deviation for the high-dose arm. Depression was also strongly

associated with dropout. These patterns of higher illness among the censored were consistent over the course of the trial. We found that applying the weights did not resolve these differences and in some instances exacerbated them. This is consistent with the weights' distribution having means farther from one among the placebo and low-dose arms (Figure 2). Concerningly, the treatment effects for the high-dose vs. placebo arms became more pronounced after weighting, from -9.1 (95%CI -14.2, -4.1) before weighting to -11.9 (95%CI -17.8, -6.0) after weighting; likewise, the treatment effect in the low-dose arm increased from -4.0 (95%CI -9.3, -1.4) to -9.2 (95%CI -15.1, -3.2). Examining the covariate distribution post-weighting informs us that the changes in effect size are likely not due to bias reduction and should not be taken seriously. For these reasons, we only present the unweighted change in PANSS score in Figure 4. The weights' poor performance is likely due to finite sample and related issues where patients who dropout lack similar counterparts among the uncensored data who can stand in for them.

Conclusions

We found the diagnostics to be instrumental in carrying out an objective analysis and were extremely useful in building and refining models for the dropout mechanism. As we expected to find, dropout was indeed informative of markers of treatment effect. As part of this project we developed freely available SAS macros that can be used to implement the visualizations reported here, available on the lead author's GitHub page. Example code for using these tools is provided in the Appendix.

References

1. Lavori PW. Clinical trials in psychiatry: should protocol deviation censor patient data? *Neuropsychopharmacology*. 1992;6(1):39-48-63.
2. Ten Have TR, Normand S-LT, Marcus SM, Brown CH, Lavori P, Duan N. Intent-to-Treat vs. Non-Intent-to-Treat Analyses under Treatment Non-Adherence in Mental Health Randomized Trials. *Psychiatr Ann*. 2008;38(12):772-783. doi:10.3928/00485713-20081201-10.
3. Wahlbeck K, Tuunainen A, Ahokas A, Leucht S. Dropout rates in randomised antipsychotic drug trials. *Psychopharmacology (Berl)*. 2001;155(3):230-233. doi:10.1007/s002130100711.
4. Kemmler G, Hummer M, Widschwendter C, Fleischhacker WW. Dropout Rates in Placebo-Controlled and Active-Control Clinical Trials of Antipsychotic Drugs. *Arch Gen Psychiatry*. 2005;62(12):1305. doi:10.1001/archpsyc.62.12.1305.
5. Liu-Seifert H, Adams DH, Kinon BJ. Discontinuation of treatment of schizophrenic patients is driven by poor symptom response: a pooled post-hoc analysis of four atypical antipsychotic drugs. *BMC Med*. 2005;3(1):21. doi:10.1186/1741-7015-3-21.
6. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615-625.
7. Hernán MA. Invited Commentary: Selection Bias Without Colliders. *Am J Epidemiol*. 2017;185(11):1048-1050.
8. Davis SM, Koch GG, Davis CE, LaVange LM. Statistical approaches to effectiveness measurement and outcome-driven re-randomizations in the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) studies. *Schizophr Bull*. 2003;29(1):73-80. doi:10.1093/oxfordjournals.schbul.a006993.
9. Kinon BJ, Liu-Seifert H, Adams DH, Citrome L. Differential Rates of Treatment Discontinuation in Clinical Trials as a Measure of Treatment Effectiveness for Olanzapine and Comparator Atypical Antipsychotics for Schizophrenia. *J Clin Psychopharmacol*. 2006;26(6):632-637. doi:10.1097/01.jcp.0000245563.06660.0f.
10. Little R, Yau L. Intent-to-Treat Analysis for Longitudinal Studies with Drop-Outs. *Biometrics*. 1996;52(4):1324-1333.
11. Fitzmaurice G, Laird NM, Ware JH. Missing Data and Dropout: Overiew of Concepts and Methods. In: *Applied Longitudinal Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Ltd; 2011:489-513.
12. Hogan JW, Laird NM. Intention-to-Treat Analyses for Incomplete Repeated Measures. *Hogan, Joseph W Laird, Nan M*. 1996;52(3):1002-1017.
13. Hernan MA, Robins JM. Longitudinal causal inference. In: *International Encyclopedia of the Social & Behavioral Sciences*. 2nd ed. Oxford, England: Elsevier; 2015:340-344.
14. VanderWeele TJ, Jackson JW, Li S. Causal inference and longitudinal data: a case study of religion and mental health. *Soc Psychiatry Psychiatr Epidemiol*. 2016;51(11):1457-1466. doi:10.1007/s00127-016-1281-9.
15. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons; 1987.
16. Robins JM, Finkelstein DM. Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests. *Biometrics*. 2000;56(3):779-788.
17. Jackson JW. Diagnostics for confounding of time-varying and other joint exposures. *Epidemiology*. 2016;27(6):1. doi:10.1097/EDE.0000000000000547.
18. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679. doi:10.1002/sim.6607.
19. Stroup TS, McEvoy JP, Swartz MS, et al. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. *Schizophr Bull*. 2003;29(1):15-31.

doi:10.1093/oxfordjournals.schbul.a006986.
20. Canuso CM, Lindenmayer JP, Kosik-Gonzalez C, et al. A randomized, double-blind, placebo-controlled study of 2 dose ranges of paliperidone extended-release in the treatment of subjects with schizoaffective disorder. *J Clin Psychiatry*. 2010;71(5):587-598. doi:10.4088/JCP.09m05564yel.
21. Hernán MA, Robins JM. *Causal Inference*. (forthcoming): CRC Press; 2017.
22. Fitzmaurice G, Laird NM, Ware JH. Missing Data and Dropout: Multiple Imputation and Weighting Methods. In: *Applied Longitudinal Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Ltd; 2011:515-550.
23. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22(3):278-295. doi:10.1177/0962280210395740.
24. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656-664. doi:10.1093/aje/kwn164.
25. Li L, Shen C, Li X, Robins JM. On weighting approaches for missing data. *Stat Methods Med Res*. 2013;22(1):14-30. doi:10.1177/0962280211403597.On.
26. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560. doi:10.1097/00001648-200009000-00011.
27. Platt RW, Schisterman EF, Cole SR. Time-modified confounding. *Am J Epidemiol*. 2009;170(6):687-694. doi:10.1093/aje/kwp175.
28. Howe CJ, Cole SR, Chmiel JS, Muñoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. *Am J Epidemiol*. 2011;173(5):569-577. doi:10.1093/aje/kwq385.
29. Shortreed SM, Laber L, Stroup TS, Pineau P, Murphy A. A multiple imputation strategy for sequential multiple assignment randomized trials. *Stat Med*. 2014;33(24):4202-4214. doi:10.1002/sim.6223.
30. Lieberman EA, Stroup TS, McEvoy JP, et al. Effectiveness of Antipsychotic Drugs in Patients with Chronic Schizophrenia. *N Engl J Med*. 2005;353(12):1209-1223. doi:10.1056/NEJMoa1412098.
31. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*. 2007;16(3):219-242.
32. van Buuren S, Oudshoorn CGM. MICE: multivariate imputation by chained equations. *R Packag version*. 2007;1(3):2007. doi:10.18637/jss.v045.i03>.
33. Seaman S, White I. Inverse Probability Weighting with Missing Predictors of Treatment Assignment or Missingness. *Commun Stat - Theory Methods*. 2014;43(16):3499-3515. doi:10.1080/03610926.2012.700371.
34. Tchetgen Tchetgen EJ, Glymour MM, Weuve J, Robins J. Specifying the Correlation Structure in Inverse-Probability-Weighting Estimation for Repeated Measures. *Epidemiology*. 2012;23(4):644-646. doi:10.1097/EDE.0b013e31825727b5.
35. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit Anal*. 2007;15(3):199-236. doi:10.1093/pan/mpl013.
36. Rubin DB. For objective causal inference, design trumps analysis. *Ann Appl Stat*. 2008;2(808-840).
37. Sun B, Tchetgen Tchetgen EJ. On Inverse Probability Weighting for Nonmonotone Missing at Random Data. *J Am Stat Assoc*. 2016. doi:10.1080/01621459.2016.1256814.
38. Shortreed SM, Moodie EEM. Estimating the optimal dynamic antipsychotic treatment regime: Evidence from the sequential multiple assignment randomized CATIE Schizophrenia Study. *J R Stat Soc Ser C Appl Stat*. 2012;61(4):577-599. doi:10.1111/j.1467-9876.2012.01041.x.
39. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-764. doi:10.1093/aje/kwv254.
40. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via

inverse probability weighting. *Basic Clin Pharmacol Toxicol*. 2006;98(3):237-242. doi:10.1111/j.1742-7843.2006.pto_329.x.

Table 1. Baseline characteristics and disposition in the paliperidone ER study			
	Placebo	Low-dose	High-dose
	N=107 (100)	N=104 (100)	N=96 (100)
Age	37.1 (10.8)	38 (9.6)	36.5 (10.3)
Gender			
Male	67 (63)	67 (64)	63 (66)
Female	40 (37)	37 (36)	33 (34)
Race/Ethnicity			
White	53 (50)	45 (43)	42 (44)
Asian	34 (32)	34 (33)	33 (34)
Black	20 (19)	23 (22)	20 (21)
Other	0 (0)	2 (2)	1 (1)
Randomization Strata			
Receiving mood stabilizer	41 (38)	38 (37)	38 (40)
Country			
Asia	33 (31)	34 (33)	30 (32)
Europe and other	35 (33)	31 (30)	31 (32)
North America	39 (36)	39 (38)	35 (36)
Baseline measures			
Total PANSS	91.6 (12.5)	96.0 (13.1)	93.3 (13.0)
CGI-Severity	4.6 (0.6)	4.6 (0.6)	4.6 (0.6)
YRMS	18.1 (6.8)	17.8 (7.4)	18.5 (8.2)
HAMD	15.7 (7.9)	17.6 (8.5)	17.3 (8.9)
Simpson-Agnes EPS	0.1 (0.2)	0.1 (0.3)	0.1 (0.3)
Weight	174.8 (44.7)	162.8 (39.7)	171.5 (49.1)
Disposition			
Remained in study	62 (63)	74 (71)	79 (76)
Ended double-blind treatment	44 (41)	34 (33)	34 (22)
… due to adverse event	7 (16)	9 (26)	4 (12)
… due to lack of efficacy	22 (50)	12 (35)	9 (26)
… due to loss to follow-up	3 (7)	3 (9)	2 (6)
… due to other reasons	0 (0)	4 (12)	1 (3)
… due to withdrawal of consent	12 (27)	6 (18)	5 (15)

Mean (sd) for continuous, N (%) for categorical variables
Acronyms: Positive and Negative Symptom Syndrome Scale (PANSS); Clinical Global Impressions Severity (CGI-Severity); Youth Mania Rating Scale (YRMS); Hamilton Depression Rating Scale (HAMD); Extrapyramidal Symptoms (EPS)
PANSS, CGI-Severity, YRMS, and HAMD were measured at target days 0, 4, 8, 15, 22, 29, 43 and when double blind treatment ended early
Simpson-Agnes EPS score was only measured at baseline and target day 15, therefore only baseline values were used. The score was taken as the mean of the tremor, fixation, elbow, shoulder, arm, and gait items to match the CATIE analysis
Percentages for reasons for ending double blind treatment sum to 100 among those who did so

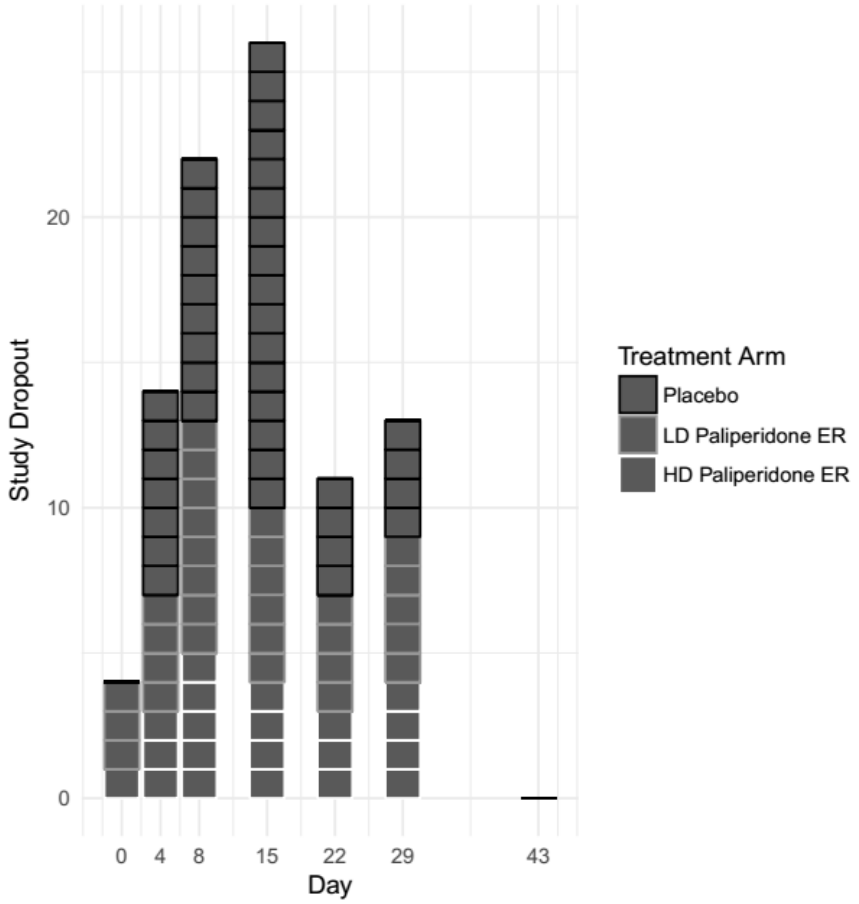

Figure 1. Study Dropout over the course of the Paliperidone ER study, by treatment arm

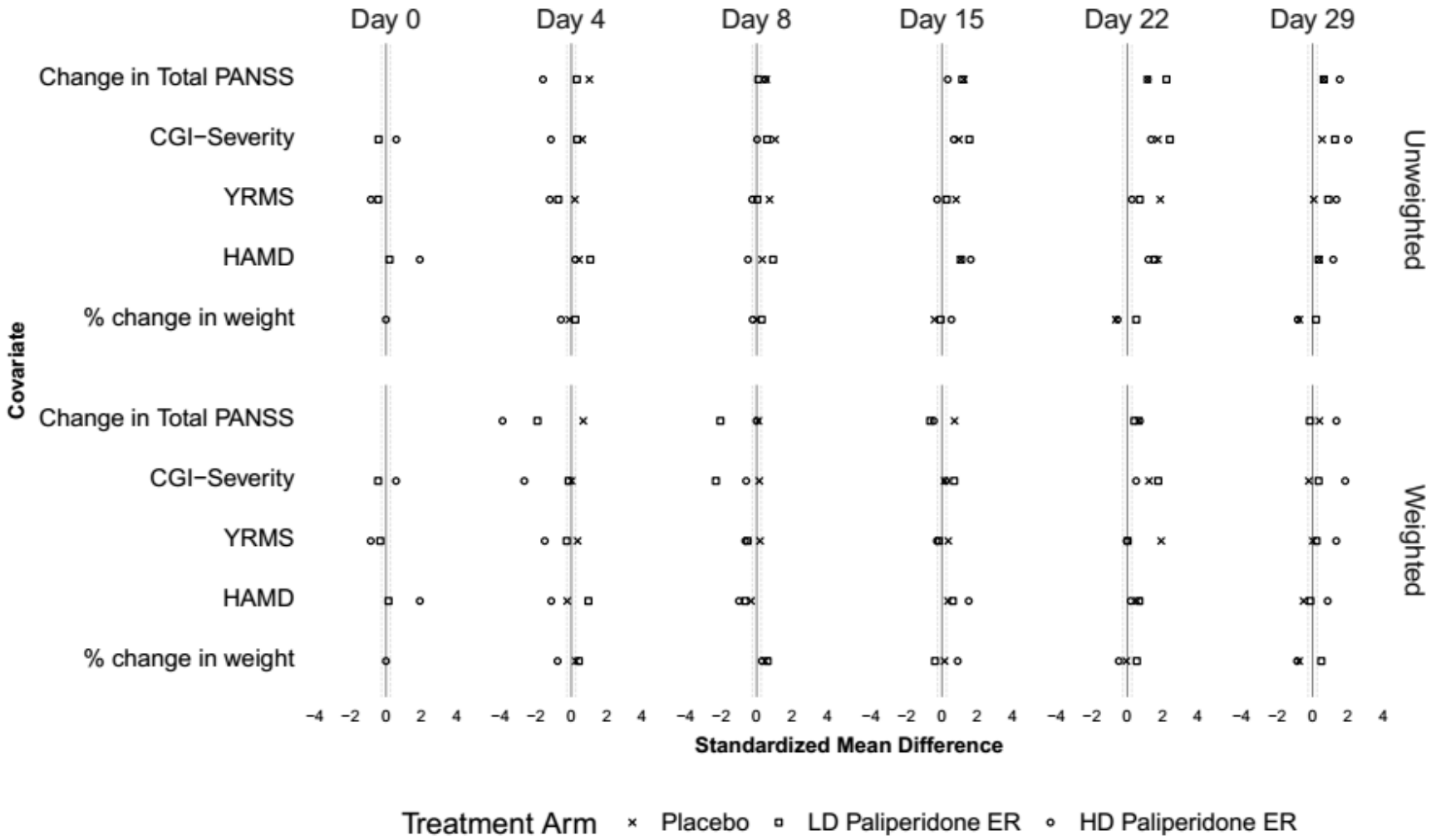

Figure 2. Mean difference of time-varying covariates comparing the censored vs. the uncensored over the course of the Paliperidone ER study, divided by the standard deviation among the uncensored, before and after application of IPCW weights. Colors distinguish the standardized mean differences for specific treatment arms. The dashed grey reference lines are placed at a positive and negative quarter of a standardized mean difference.

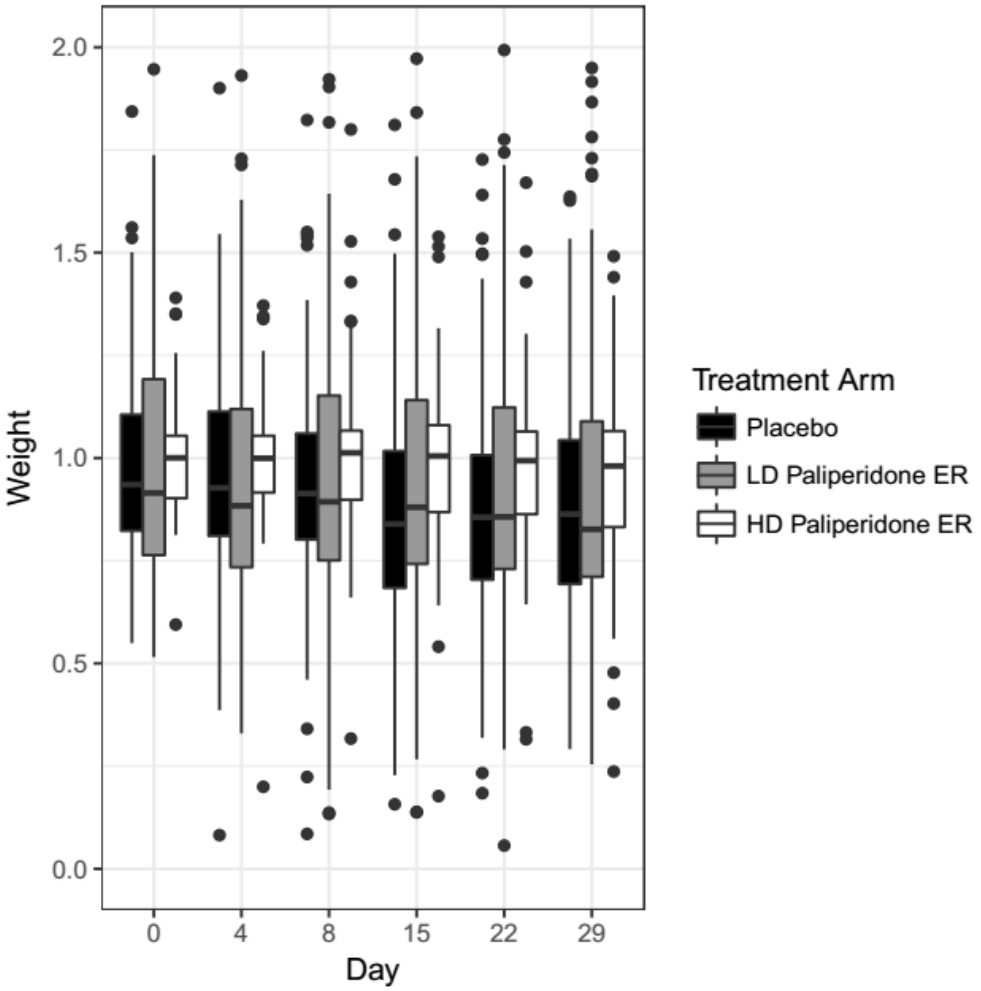

Figure 3. Boxplots depicting the distribution of the stabilized IPCW weights over the course of the study, by treatment arm. Though not shown, the full range of the weights was [0.06, 32.4] indicating very influential observations. The weights for the low-dose and placebo arms substantially diverge from one over time.

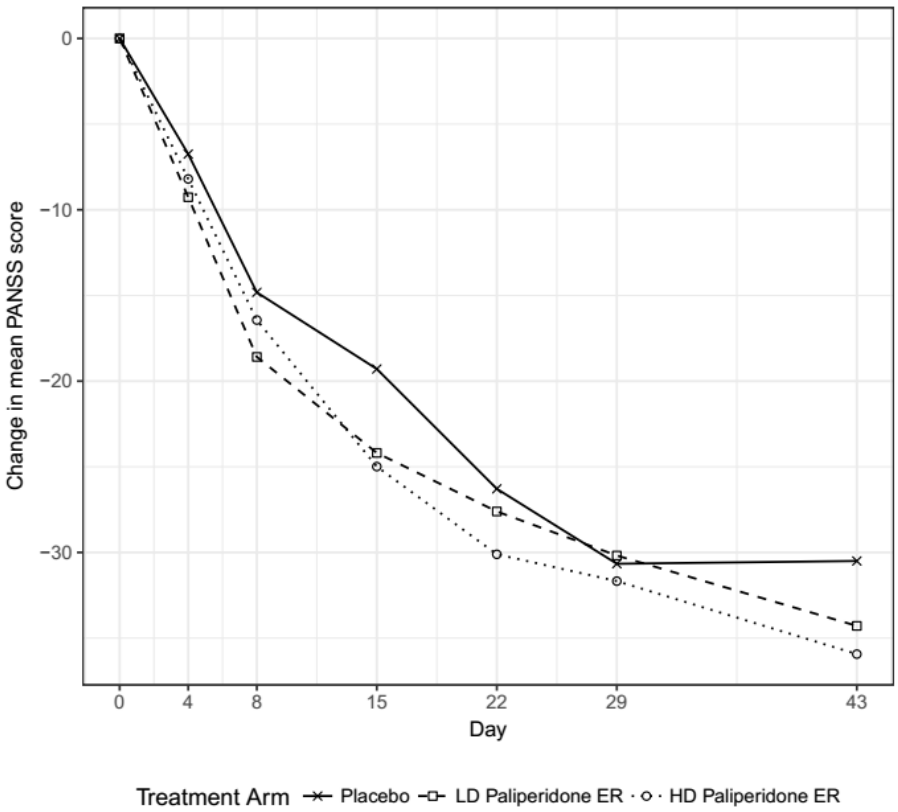

Figure 4. Change in Total PANSS score over the course of the Paliperidone ER study, by treatment arm.

Appendix

Weight specification

$$W(t) = \frac{\mathrm{P}[A(0) = a|V]}{\mathrm{P}[A(0) = a|C(0), V]} \prod_{m=0}^{t} \frac{\mathrm{P}[S(m+1) = 0|\bar{S}(m) = 0, A, V]}{\mathrm{P}[S(m+1) = 0|C(m), \bar{S}(m) = 0, A, V]}$$

where t represents a specific follow-up visit, m represents a visit equal to or less than t, $S(m+1)$ represents dropping out of the study immediately after visit m (0 = remains at visit $m+1$, 1 = drops out by visit $m+1$), $\bar{S}(m) = 0$ represents no censuring up through month m, $C(m)$ represents covariates measured at visit m, while A and V respectively represent assigned treatment and covariates at baseline.

Example SAS code

Note: The following SAS macros are available at the lead author's faculty page through a link to a GitHub repository. A software manual detailing the functions is provided, along with a toy dataset.

```
%lengthen(input=YodaWide,
output=YodaLong,
save=MyOtherDatasets,
diagnostic=3,
censoring=no,
id=yodaid,
exposure=studydisc,
times_exposure=0 4 8 15 22 29,
times_covariate=0 4 8 15 22 29,
weight_exposure=cumw,
temporal_covariate = Chgpansstotal hamdtotal yrmstotal cgisev pctchgweight,
history=treatgrp);

%balance(input=YodaLong,
save=YodaWide MyOtherDatasets,
output=YodaTable,
diagnostic=3,
approach=weight,
censoring=no,
scope=recent,
recency=0,
exposure=studydisc,
history=treatgrp,
weight_exposure=cumw,
times_exposure = 0 4 8 15 22 29,
times_covariate = 0 4 8 15 22 29,
sort_order=alphabetical);

%makeplot(input=YodaTable,
save=YodaLong YodaWide MyOtherDatasets,
Output=YodaPlot,
approach=weight,
metric=SMD,
scope=recent)
```