Doubly Robust Methods for Analysis of RCTs

Tim Feeney

10/2/2021

COI, Funding, Disclaimer.

- research editor at The BMJ
- consulting for iHeed Medical Education
- research funded by R01AI157758
- all errors are mine

Intention to Treat (ITT)

- ITT estimators are unbiased in randomized trials
- Robust property derives from leveraging randomization to assure exchangeability in expectation (in addition to other identifiability conditions)
- ITT is often imprecise due to lack of covariate predictor adjustment
- However, even with covariate adjustment, strong assumptions are made about the relationship between covariates and the outcome

Can we do more?

Background of the AIPW estimator

- 1994 JASA Robins, Rotnizky and Zhao
- 2006 Tsiatis 'Semiparametric Theory and Missing Data'

$$\hat{\mu}_1 = n_1^{-1} \sum_{i=1}^{n_1} \left[\frac{R_i Y_{1i}}{\pi\left(W_i, \hat{\gamma}_3\right)} - \frac{\left\{R_i - \pi\left(W_i, \hat{\gamma}_3\right)\right\}}{\pi\left(W_i, \hat{\gamma}_3\right)} \mu\left(W_i, \hat{\gamma}_1\right) \right]$$

- 2011 AJE Jonsson-Funk, Westreich, Wiesen, Sturmer, Brookhart and Davidian 'Doubly Robust Estimation of Causal Effects'

$$\begin{aligned}\hat{\Delta}_{\mathrm{DR}} =& n^{-1} \sum_{i=1}^{n} \left[\frac{X_i Y_i}{e\left(\mathbf{Z}_i, \hat{\beta}\right)} - \frac{\left\{X_i - e\left(\mathbf{Z}_i, \hat{\beta}\right)\right\}}{e\left(\mathbf{Z}_i, \hat{\beta}\right)} m_1\left(\mathbf{Z}_i, \hat{\alpha}_1\right) \right] \\ & - n^{-1} \sum_{i=1}^{n} \left[\frac{\left(1 - X_i\right) Y_i}{1 - e\left(\mathbf{Z}_i, \hat{\beta}\right)} + \frac{\left\{X_i - e\left(\mathbf{Z}_i, \hat{\beta}\right)\right\}}{1 - e\left(\mathbf{Z}_i, \hat{\beta}\right)} m_0\left(\mathbf{Z}_i, \hat{\alpha}_0\right) \right]\end{aligned}$$

Key Points of AIPW for ITT

- Consistent(asymptotically unbiased) under MAR and positivity even if the one of the models is wrong [1]
- Added bonus: If both models are right then this is the most efficient estimator in its class [1]

Under RCT we are guaranteed positivity and exchangeability, and we KNOW one model is right and WHICH model is right, too.

[1] 1 Daniel RM. Double Robustness. In: Wiley StatsRef: Statistics Reference Online. 2018. doi:10.1002/9781118445112.stat08068

Simulation/ Proof of concept-Binary

Binary outcome

10000 monte carlo trials

$n = 1000$

$Y \sim$ binomial

weak, med, strong predictor, error, non predictor $\sim N(0, 1)$

	Estimate	SE	se_ratio
ITT	0.2781	0.0286	1.0000
Non Predictors	0.2509	0.0273	0.9548
Weak Predictors	0.2514	0.0274	0.9579
Strong Predictor	0.2410	0.0217	0.7598
All Predictors	0.2628	0.0198	0.6936

Simulation/ Proof of concept-Continuous

Continuous outcome

10000 monte carlo trials

$n = 1000$

$Y \sim continuous$

	Estimate	SE	se_ratio
ITT	4.2836	0.7228	1.0000
null Predictor	4.2836	0.7148	0.9889
non Predictor	4.2838	0.7145	0.9885
Weak Predictor	4.2939	0.7176	0.9929
Medium Predictor	4.1819	0.6606	0.9139
Strong Predictor	5.1438	0.3450	0.4773
All Predictors	5.0600	0.1256	0.1738

Real Data

Reanalysis of ACTG 5202 data based off Sax 2009 NEJM [2]

- Phase 3 clinical trial
- Comparing abacavir/lamivudine (ABC/3TC) to emtricitibine/tenofivir (TDF/FTC)
- main outcome was time to virologic failure
- 797 person randomized had $\geq 100,000$ HIV RNA copies/ml, were $\geq$ 16years old and had at most 7 days of ART therapy
- primary result was hazard ratio of 2.33 (1.46-3.72) for ABC/3TC versus TDF/FTC

[2] 1 Sax PE et al. Abacavir–Lamivudine versus Tenofovir–Emtricitabine for Initial HIV-1 Therapy. N Engl J Med 2009;doi:10.1056/NEJMoa0906768

Real Data-Slighty modified

In this analysis we are approaching categorical and continuous outcome–we have not pursued survival analysis,yet.

- we used data at 48 weeks to define viral supression defined as HIV RNA copies $<$ 200 copies/ml –>our binary outcome
- similarly we used data at 48 weeks to define CD4 count, cells/mm^2 –> our continuous outcome
- Use the g-formula as out outcome model; others models are potential candidates, too
  - predictors included in g-formula are
    - Age(categorized, 16-24, 25-49, 50+)
    - log RNA level at baseline
    - CD4 count at baseline
    - Sex at birth

Confidence Intervals

- Bootstrapping to obtain a standard error; can also used closed form solution [3]

Also a closed form solution available:

$$\widehat{\mathbb{V}}\left(\widehat{\mu}_{aipw}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\left[\frac{R_i Y_i}{\pi\left(\mathbf{X}_i;\widehat{\boldsymbol{\alpha}}\right)} + \left\{1 - \frac{R_i}{\pi\left(\mathbf{X}_i;\widehat{\boldsymbol{\alpha}}\right)}\right\} m\left(\mathbf{X}_i;\widehat{\boldsymbol{\beta}}\right) - \widehat{\mu}_{aipw}\right]^2$$

R code, which I think is a bit easier to read

```
aipw_diff<-(aipw1-aipw0)-aipw_all_est

var_aipw<-(1/(nrow(data)^2))*sum(aipw_diff^2)
se_aipw<-sqrt(var_aipw)
```

[3] Learned closed form solution thanks to Paul Zivich

Reanalysis results

	Viral suppression at 48 weeks	SE	SE Ratio
ITT	-0.025 (-0.065,0.016)	0.0205	ref
AIPW	-0.020 (-0.019,0.061)	0.0203	0.99

	CD4 count at 48 weeks (cells/mm^2)	SE	SE Ratio
ITT	-6.85 (-38.92,25.23)	16.34	ref
AIPW	-6.69 (-26.94,13.57)	10.33	0.63

Implications

- Tighter confidence intervals from lower SE around unbiased estimates
- Efficient use of valuable data
- Sample size gains!

Sample Size

RCTs are typically designed with power in mind to make sure that the correct number of people are enrolled to answer a question without enrolling too many people as to make the trial unethical.

- Power typically set at 80-90%

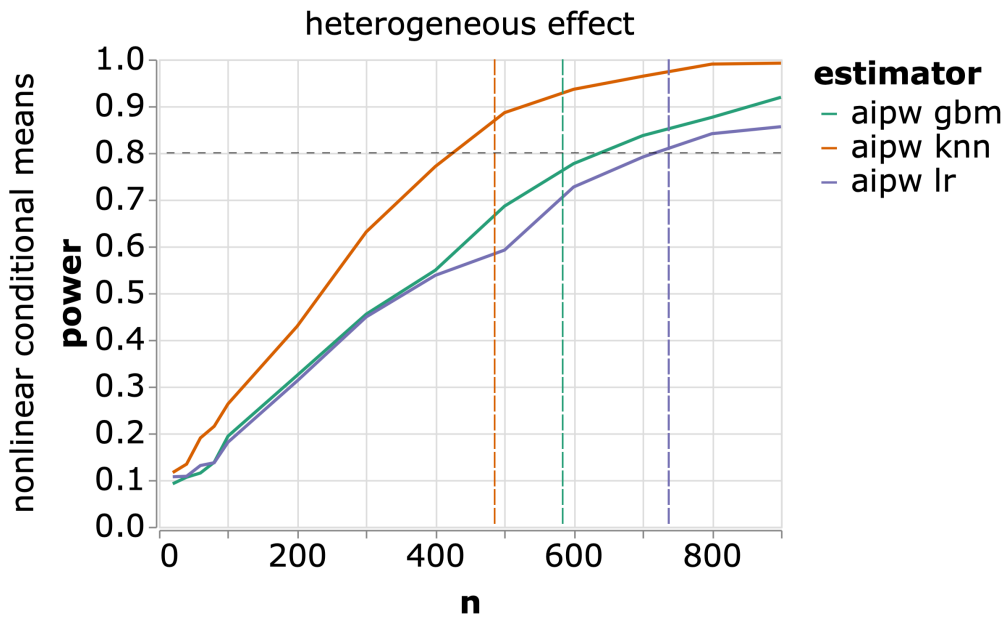
Using AIPW and leveraging the efficiency gains (as evident by the decrease in SE) means you can use less people to answer the same question.

Alternatively, for a study that failed to recruit the appropriate number of participants, AIPW might be a way to recover power

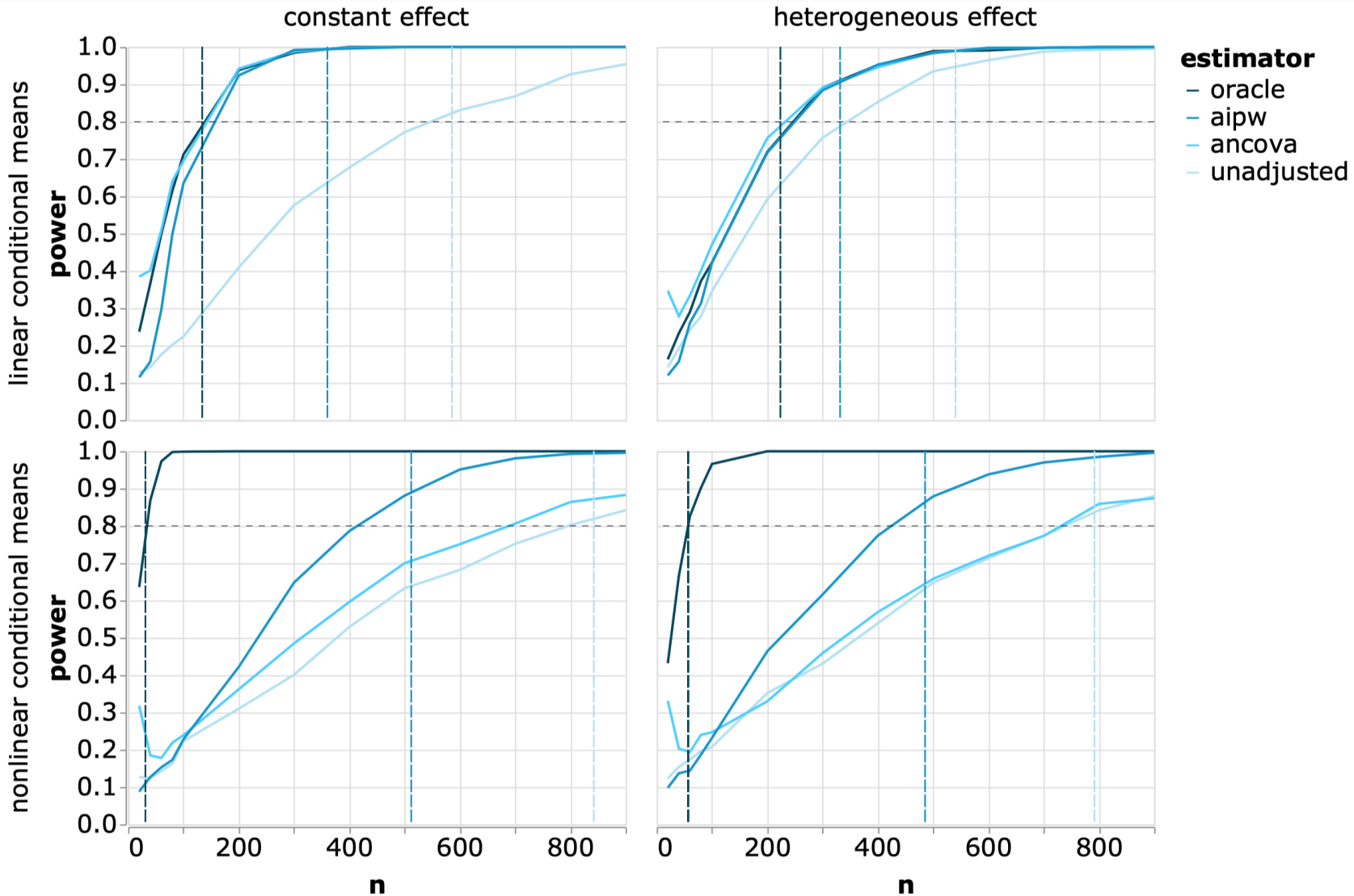
Sample size improvements

As far as I know there are no closed form solutions for calculating sample size 'savings'

Simulation seems to be the best solution [4], which requires many assumptions about the covariance structure of the data



[4] Schuler A. Designing efficient randomized trials: power and sample size calculation when using semiparametric efficient estimators. arXiv:210410784. 2021.http://arxiv.org/abs/2104.10784


constant effect
heterogeneous effect
linear conditional means
nonlinear conditional means
power
n
estimator
oracle
aipw
ancova
unadjusted
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
200
400
600
800

Finishing up

As I finish this project I want to make sure I quantify the sample size improvements we should be able to see.

Plan is to evaluate the number of observations I would need to add to my data to achieve the SE of the ITT estimate.

Thank you

Many awesome people helped me with this project, taught me, and tolerated code review sessions.

- Steve Cole
- Jess Edwards
- Paul Zivich
- Rachael Ross
- Catherine Li
- Lauren Zalla
- Audrey Renson