Improving Efficiency of Inferences in Randomized Clinical Trials Using Auxiliary Covariates

Min Zhang, Anastasios A. Tsiatis and Marie Davidian
(2008, Biometrics)

Presented by Rui Zhang

May 31, 2012

A motivating example

- Suppose in a randomized clinical trial (RCT), whether or not getting a certain disease after treatment (treated by drug or placebo, denoted by $Z = 1, 0$) is recorded in every patient as Y, and also baseline covariates, such as demographic information, is also recorded *before treatment assigned* as X, then how do we compare the treatment effect?
- One quick answer:

$$\mathit{logit}\{\mathbb{E}(Y|Z)\} = \beta_0 + \beta_1(Z = 1)$$

- How can we incorporate information from X?

One quick answer: adjustment

- Linear links – good idea!

  No bias, gain in efficiency.

- Non-linear – problematic!

		Un-adjusted	Adjusted
Linear Model	MC bias of $\hat{\beta}'$	0.001	0.000
$\beta = 2.000$	MC mean in $s.e.(\hat{\beta}')$	0.04161	0.008160
Logistic Model	MC bias of $\hat{\beta}'$	0.014	0.108
$\beta = 1.910$	MC mean in $s.e.(\hat{\beta}')$	0.2089	0.2165

Adjustment consequence in *logistic model:*

1. β further away from 0 and s.e. larger;
2. Interpretation is subgroup-specific, not marginal.

A quote

"Every clinical trial is a problem of missing data."

– Professor Scott Emerson

Missing data problem?

- Start with 2-treatment arm scenario;
- Suppose outcomes are (Y_0, Y_1) for each patient under placebo and treatment;
- Treatment group: $Y_1's$ observed and Y_0 missing
- Control group: $Y_0's$ observed and Y_1 missing
- $\Rightarrow$ **Missing at random**
- Moreover, in RCT, it is **Missing completely at random**

Missing data strategy: estimating $\mu_1 = \mathbb{E}Y_1$

- Complete case analysis: use observed Y_1's average.

$$\hat{\mu}_{11} = n_1^{-1} \sum_{i=1}^{n} Z_i Y_{1i}$$

- Inverse weight method:
  We utilize information in X by estimating probability of non-missed given some X:

$$\pi(x) \equiv Pr(Z = 1 | X = x)$$

  and thus re-weighted average of Y_1 becomes:

$$\hat{\mu}_{12} = n_1^{-1} \sum_{i=1}^{n} \frac{Z_i Y_{1i}}{\hat{\pi}(X_i)}$$

μ_1 *estimation continued*

- Imputation by regression:
  We utilize information in X by estimating conditional expectation of Y_1 given $X's$, $\mathbb{E}(Y_1|X)$, in the form as $X\gamma_1$:

  $$\hat{\mu}_{13} = n_1^{-1}\sum_{i=1}^{n} x_i\hat{\gamma}_1$$

  where $\hat{\gamma}_1$ is obtained via regression observed Y_1's on their respective $X's$.
- Double Robust Estimator: Combining $\pi(X)$ and $\mathbb{E}(Y_1|X)$:

  $$\hat{\mu}_{14} = n_1^{-1}\sum_{i=1}^{n}\left(\frac{Z_i}{\hat{\pi}(X_i)}(Y_{1i} - X_i\hat{\gamma}_1) + X_i\hat{\gamma}_1\right)$$

Double-Robust Estimator

- Double-robustness: whichever γ_1 or $\pi(X_i)$, is modelled correctly, the estimator is consistent.
- In RCT: $\pi(X_i)$ is always correct! It is a constant.
- More generally, replacing
  - Y_i with U_i, the marginal x_i-free estimating equation;
  - $\mu(X_i, \hat{\gamma}_1)$ with ϕ_i, some arbitrary term that may depend on x_i;

  and denote $\frac{Z_i}{\pi(X_i)}$ as w_i and in RCT, it is a constant. We have

$$\sum_{i=1}^{n} (w_i U_i + (1 - w_i)\phi_i) = 0$$

- Choosing $\phi_i = \mathbb{E}(U_i | x_i)$ gives a semi-parametric efficient estimator.

Our more general "Double-Robust Estimator": from an "augmented" EE

- In this paper, we proposed a more general estimator by considering multiple treatment groups.
-

$$U(Y, Z; \beta) + \sum_{g=1}^{k} (I(Z = g) - \pi_g)\, \mathbb{E}(U(Y, Z; \beta)|X_i, Z_i = g) = 0 \tag{1}$$

- First part: original estimating equation;
- Second part: "*augmented part*";
- It reaches semi-parametric efficiency bound.

A small summary

$$U(Y, Z; \beta) + \sum_{g=1}^{k} (I(Z = g) - \pi_g)\, \mathbb{E}(U(Y, Z; \beta)|X_i, Z_i = g) = 0$$

- By double-robust estimator property, our estimator should always be consistent–beating adjustment estimators!
- By *"augmented part"*, new EE has lower variance and thus new estimator has lower variance–beating original un-adjustment estimators!
- Estimator from above EE conveniently has Wald test and consequently a (locally most) power χ^2 test can be derived!

How do we get to it?

- Instead of generalizing original double robust estimators proposed by other papers and show it is semi-parametric efficient, this paper directly works from a semi-parametric model.
- Proposing their semi-parametric model.
  - Step 1
    Consider Y and Z jointly as we did in "classic" analysis
    $$f(y, z; \beta, \eta)$$
    β is parameter of interest: (μ_0, μ_1)
    η is nuisance parameter. For example
    - Y is normal, then η is σ, the standard deviation;
    - Y is binary, then no η;
  - Step 2
    Consider Y, Z and X jointly
    $$f(y, z, x; \beta, \eta, \psi)$$
    ψ a nuisance parameter to parametrize the three jointly.

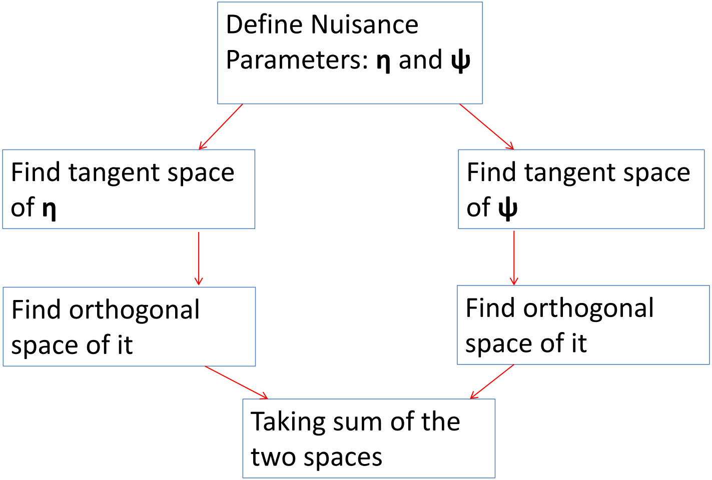
How do we get to it? – Work by parametric sub-model

- But we do not want to assume anything further than β so η and ψ are going to be anything the data tell us $\rightarrow$ infinite-dimensional parameters
- Consider a subset of the semi-parametric model where (η, ψ) are finite-dimensional and contain the truth (η_0, ψ_0):

$$\eta = \eta_0 + \alpha u(\alpha) \quad \psi = \psi_0 + \tau v(\tau)$$

  where $u(\cdot)$ and $v(\cdot)$ are fixed functions.
- When $\alpha = \tau = 0$, we have the truth.
- Now we have a set of parametric models!
- With nuisance parameters η and ψ, the most efficient estimator of β should be orthogonal to nuisance tangent space, which is spanned by likelihood scores of η and ψ

Working streamline for a parametric model



Work in finite and generalize into infinite

- The left part is ready: our original EE;
- The right part is complicated to work with but we can show it is
$$\{h(X, Z) : \mathbb{E}\{h(X, Z)|X\} = 0\}$$
- Combining them together we get:

$$U(Y, Z; \beta) + \sum_{g=1}^{k} (I(Z = g) - \pi_g)\, \mathbb{E}(U(Y, Z; \beta)|X_i, Z_i = g) = 0$$

- Since $u(\cdot)$ and $v(\cdot)$ are any fixed functions, our EE should work for all parametric models contained in the semi-parametric model
- The above EE works for semi-parametric model!

Simulation: binary outcome Y

- A binary outcome from a two-arm RCT of 600 subjects, with $5,000$ Monte Carlo datasets.
- Data generating mechanism:

$$logit\,(Pr(Y=1|Z=g,X)) = \alpha_{0g} + \alpha_g^T X \quad g=1,2$$

- Correlation levels between Y and X

	$\frac{Var(\mathbb{E}(Y_1\|X))}{Var(Y_1)}$	$\frac{Var(\mathbb{E}(Y_0\|X))}{Var(Y_0)}$
Mild association	0.16	0.18
Moderate association	0.33	0.32
Strong association	0.8	0.38

- Original un-adjusted estimator;
- Proposed estimator from "*augmented*" EE;
- Adjusted for $X's$ estimator.

Simulation results

Method	β_2	MC Bias	MC SD	Ave. SE	Cov. Prob	Rel. Eff.
			Mild Association			
Unadjusted	-0.494	-0.00044	0.1668	0.1661	95.0%	1.00
Aug.	-0.494	-0.00042	0.1545	0.1533	94.9%	1.17
Adjusted	-0.494	-0.091	0.1831	0.1822	92.6%	0.67
			Moderate Association			
Unadjusted	-0.490	-0.0025	0.1634	0.1650	95.5%	1.00
Aug.	-0.490	-0.0026	0.1390	0.1392	95.1%	1.38
Adjusted	-0.490	-0.2208	0.2015	0.2015	81.2%	0.30
			Strong Association			
Unadjusted	-0.460	-0.0026	0.1662	0.1655	95.2%	1.00
Aug.	-0.460	-0.0026	0.132	0.131	95.2%	1.59
Adjusted	-0.460	-0.3266	0.222	0.2210	68.8%	0.18

Simulation: more powerful test

- A normal outcome from a three-arm RCT of 200 or 400 subjects, with $10,000$ Monte Carlo datasets:

$$\begin{pmatrix} Y \\ X \end{pmatrix} | Z \sim \mathcal{N}\left(\begin{pmatrix} \beta_1 I(Z=1) + \beta_2 I(Z=2) \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

  in which β is the parameter being estimated.
- Mild association: $\rho = 0.25$
- Moderate association: $\rho = 0.5$
- Strong association: $\rho = 0.75$
- Null hypothesis: $\beta_1 = \beta_2 = 0$
- Alternative hypothesis: $\beta_1 = 0.25$, $\beta_2 = 0.4$
- Compared to: Kruskal-Wallis Test (reduces to Wilcoxon-rank test when $k = 2$).

Simulation Results

		Null		Alternative	
ρ	n	K-W Test	Our Test	K-W Test	Our Test
0.25	200	0.05	0.05	0.51	0.57
	400	0.05	0.05	0.83	0.87
0.50	200	0.05	0.05	0.51	0.67
	400	0.05	0.05	0.83	0.94
0.75	200	0.05	0.05	0.50	0.89
	400	0.05	0.05	0.83	1.00

Conclusion and Discussion

$$U(Y, Z; \beta) + \sum_{g=1}^{k} (I(Z = g) - \pi_g)\, \mathbb{E}(U(Y, Z; \beta)|X_i, Z_i = g) = 0$$

- By double-robust estimator property, our estimator should always be consistent–beating adjustment estimators!
- By residual bias correction, the second term, we are beating original un-adjustment estimators!
- Estimator from above EE conveniently has Wald test and consequently a (locally most) power χ^2 test can be derived!
- We are making few assumptions: no assumptions except for β;
- But...
  sometimes it can be very hard to estimate $\mathbb{E}(U(Y, Z; \beta)|X_i, Z_i = g)$ and if we are unlucky and find a wrong model to do regression, we might not gain much efficiency.

Wrong model for $\mathbb{E}(U(Y,Z;\beta)|X_i, Z_i = g)$

- Correct model:
  $\mathbb{E}(U(Y,Z;\beta)|X_i, Z_i = g) = aX_1 + bX_2 + cX_3 + dX_4$
- Regression model:
  $\mathbb{E}(U(Y,Z;\beta)|X_i, Z_i = g) \sim a^* X_1^3 + b^* e^{X_2} + c^* log(abs(X_3))$

Method	β_2	MC Bias	MC SD	Ave. SE	Cov. Prob	Rel. Eff.
			Mild Association			
Unadjusted	-0.494	-0.00044	0.1668	0.1661	95.0%	1.00
Augmented	-0.494	-0.00061	0.1598	0.1587	95.1%	1.09
			Moderate Association			
Unadjusted	-0.490	-0.0025	0.1634	0.1650	95.5%	1.00
Augmented	-0.490	-0.0029	0.1525	0.1517	95.2%	1.15
			Strong Association			
Unadjusted	-0.460	-0.0026	0.1662	0.1655	95.2%	1.00
Augmented	-0.460	-0.0050	0.1516	0.1500	95.3%	1.20

Questions?