Inference in response-adaptive trials when the patient population varies during time

Massimiliano Russo, Steffen Ventz, Victoria Wang & Lorenzo Trippa
m_russo@hms.harvard.edu

A critical assumption in CT

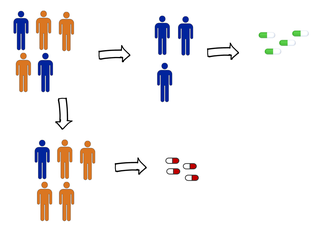
- Clinical trial (CT) analyses assume that **patient characteristics** & treatment effects are **constant** during time.

- In long trials - lasting several years - 'events' can modify the target populations.

Characteristics can vary over time & we need to account for this in the data analysis.

Causes

Example: While enrolling patients a concurrent trial targeting a sub-group of patients start enrolling patients as well.

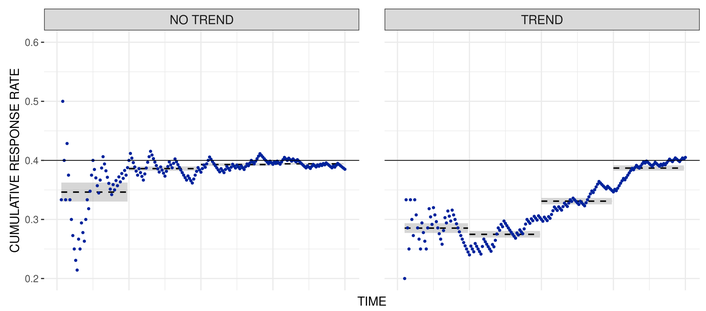


Another example is offered by the recent pandemic of COVID-19.

Consequences

- Standard techniques for testing and estimation do not account for these changes.
- Changes in patients characteristics can **bias treatment effect** estimates and **inflate type I error** rates of standard testing procedures.

Effect on trial data



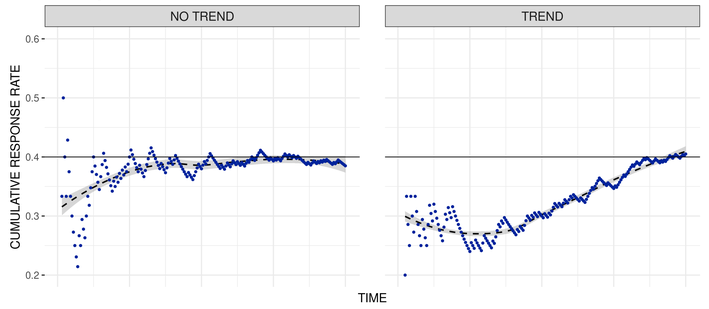
Variation in patients characteristics over time leads to trends in the response rate.

References

- Russo, M., et al., Inference in response-adaptive clinical trials when the enrolled population varies over time. (submitted)
- Jiang, F., et al., Robust alternatives to ANCOVA for estimating the treatment effect via a randomized comparative study, JASA, 2018.
- Wood, S., Thin plate regression splines, JRSS-B, 2003,

Solutions?

If **all** varying patient characteristics are observed → conditioning.
Else we can try to estimate trends.



Method 1: GAMs

We leverage flexible **Generalized Additive Model** to estimate trends of unknown patient characteristics

$$\pi_a(t,x) = F\Big(\theta_0 + \gamma_a 1\{a>0\} + \mathbf{f(t)} + \xi x\Big),$$

- $\mathbf{f(t)}$ is an unknown function of the enrollment time t.

The main advantage is **parsimony** → treatment effect can be tested using γ_a via:

$$H_{0,a}: \gamma_a \leq 0, \text{ v.s. } H_{1,a}: \gamma_a > 0,$$

rather than functionals of $\pi_a(t,x)$.
We choose f as a smoothing spline, estimating the model via penalized log-likelihood.

Method 2: A biased adjusted test procedure

We start from $\widehat{\gamma}_a$ & 'adjust' its value:

$$\widehat{\gamma}_a^{\text{adj}} = \widehat{\gamma}_a - \widehat{\mathbb{E}[\widehat{\gamma}_a - \gamma_a \mid \Delta_a]},$$

- $\Delta_a = \mathrm{d}(\boldsymbol{T}_a, \boldsymbol{T}_0)$ that quantifies the unbalance between the empirical distributions of enrollment times.
- We use $\Delta_a = \overline{T}_a - \overline{T}_0$ as proxy for the imbalance of the patient profiles.

A robust test is

$$H_{0a}: \gamma_a^{\text{adj}} \leq 0 \quad \text{v.s} \quad H_{1a}: \gamma_a^{\text{adj}} > 0,$$

$$\hat{\gamma}_a^{\text{adj}} = \{\hat{\gamma}_a - \hat{\Sigma}_{12}\hat{\Sigma}_{22}^{-1}\hat{\Delta}_a\}/\{\widehat{\Sigma}_{11} - \widehat{\Sigma}_{12}\widehat{\Sigma}_{22}^{-1}\widehat{\Sigma}_{21}\}^{1/2} \sim N(0,1).$$

- Asymptotic distribution is derived via CLT + multivariate delta method.
- If $\Delta_a \approx 0$: $\widehat{\gamma}_a$ remains in-adjusted (standard procedure).
- Δ_a can quantify the effect of other covariates beside time.

Inference

How do we do inference under adaptive design?

- Bootstrap procedures can be reliable in a variety of designs - still rely on some assumptions!
- Asymptotic approximations depending on the considered design.

Key property of the design: allocation probability for each arm $p_a > 0$, i.e. asymptotically we keep accumulating information for each arm.

Example: Glioblastoma study

- Y_i: 12 month survival; two datasets $N = 150$ trial & 320 external controls.
- Treatement: temozolomide in combination with radiation therapy (TMZ+RT) - the current SOC.
- The datasets include patient-level data: age, gender, Karnofsky performance status (KPS), and extent of tumor resection.

Generate the studies

Enrolled Population: select a patient record (Y_i, X_i)

$$Pr\Big((Y_i,X_i) = (y_j^{(0)}, x_j^{(0)}) | T_i = t\Big) \propto p_1\Big(x_{j,1}^{(0)}, t\Big) \times p_2\Big(x_{j,2}^{(0)}, t\Big)$$

$x_{j,\ell}^{(0)} \in \{0,1\}, \ell = 1,2$: patient's gender & KPS (1 if ≥ 80),

Induces a trend $t \to \pi_a(t)$: KPS and gender correlate with survival.

Randomization: We assigned treatments, $A_i \sim p(A_i | D_{1:i-1})$, according to a multi-arm BAR.

Effective arms: To introduce positive effects we randomly relabel negative outcomes $(A_i = 1, Y_i = 0)$ on arm 1 into positive outcomes $(A_i = 1, Y_i = 1)$.

Some results

Two sets of analysis:

1. We simulated the GBM trial - all patient characteristics available & used for analysis.
2. In the second setting, the KPS is not available for data analysis.

We show results from 5,000 replicates of the described study, with target $\alpha = 0.10$

No Covariates

Method	Arm 1	Arm 2
Z-Test	0.908	**0.230**
L-Test	0.587	0.083
A-GAM	0.588	0.107
B-GAM	0.600	0.096

With KPS

Method	Arm 1	Arm 2
Adj-Z-X	0.638	0.120
A-GAM-X	0.629	0.121
B-GAM-X	0.618	0.107
A-GAM-X-E	0.640	0.102
B-GAM-X-E	0.645	0.099

- Proposed correction: right type I error.
- Covariates and External data increase power.

Without KPS

Method	Arm 1	Arm 2
Adj-Z-X	0.604	0.099
A-GAM-X	0.610	0.112
B-GAM-X	0.601	0.099
A-GAM-X-E	0.620	0.100
B-GAM-X-E	0.622	0.096

Conclusions

We proposed:

- Testing procedures 'robust' to variations of patient characteristics over time.
- Combined the proposed framework with known confounder corrections.
- + External Data
- We focused on BAR **but** applicable to a broad class of adaptive designs.

Important: no loss of power when $\mathbf{f(t)} \approx 0$.

Harvard-MIT Center for Regulatory Science
HARVARD MEDICAL SCHOOL
Dana-Farber Cancer Institute