BIOMETRICS 0, 1–2 DOI: 0000-0000-0000
January 2019

Longitudinal surrogate marker analysis

Denis Agniel

RAND Corporation

email: dagniel@rand.org

and

Layla Parast

RAND Corporation

email: parast@rand.org

SUMMARY: The text of your summary. Should not exceed 225 words.

KEY WORDS: longitudinal data; surrogate markers; nonparametric analysis.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

2. Method

2.1 *Setup and notation*

Let the data for analysis consist of n independent observations of the form $(Y_i, \mathbf{X}_i, A_i)_{i=1,\ldots,n}$, A_i represents an indicator for treatment or intervention, $\mathbf{X}_i = (X_{ij})_{j=1,\ldots n_i}$ is a longitudinally collected surrogate marker, and Y_i is a primary outcome of interest, all for subject i. We assume for simplicity of presentation that patients are randomly assigned at baseline to treatment or control and that Y is fully observed. We further assume that there exists $X(\cdot)$ an underlying surrogate marker trajectory, which we only observe n_i times, possibly at only a few, irregularly spaced times and with error.

Furthermore, let $Y_i^{(1)}$ and $Y_i^{(0)}$ denote the primary outcome one would observe if, possibly contrary to fact, subject i received treatment and control, respectively. We assume the stable unit treatment value assumption (SUTVA, Rosenbaum and Rubin (1983)). Similarly, let $X_i^{(1)}$ and $X_i^{(0)}$ denote the summary markers under treatment and control. We assume that the joint distribution of Y_i and $\mathbf{X}_i$ is given by $f_j(y, \mathbf{x}) = f_j(y|\mathbf{x})g_j(\mathbf{x})$ in treatment group j where $f_j(y|\mathbf{x})$ is the density of Y conditional on $\mathbf{X} = \mathbf{x}$ and $g_j(\mathbf{x})$ is the density function for $\mathbf{X}_i$ in group $D = j$.

2.2 *Estimating treatment effects and surrogacy*

We are interested in estimating the proportion of treatment effect on the primary outcome that is explained by the longitudinal surrogate marker. We define the overall treatment effect, Δ, as the expected difference in Y under treatment and control,

$$\Delta = E(Y^{(1)} - Y^{(0)}).$$

Because of randomization, we can use the observed data to estimate Δ

$$E[Y|A = 1] - E[Y|A = 0] = \int y f_0(y|\mathbf{x})g_0(\mathbf{x})dyd\mathbf{x} - \int y f_1(y|\mathbf{x})g_1(\mathbf{x})dyd\mathbf{x}.$$

We aim to measure the surrogate value of $\mathbf{X}$ comparing Δ to the residual treatment effect that would be observed if the $\mathbf{X}$ was distributed the same in both groups. The residual treatment effect can be estimated as

$$\begin{aligned}\Delta_S &= \int E[Y|A=1,\mathbf{X}=\mathbf{x}]g_0(\mathbf{x})d\mathbf{x} - \int E[Y|A=0,\mathbf{X}=\mathbf{x}]g_0(\mathbf{x})d\mathbf{x} \\ &= \int y f_1(y|\mathbf{x})g_0(\mathbf{x})dyd\mathbf{x} - \int y f_0(y|\mathbf{x})g_0(\mathbf{x})dyd\mathbf{x}, .\end{aligned}$$

3. Simulation studies

4. Analysis of longitudinal CD4 count surrogacy

5. Discussion

References

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70,** 41–55.

Received Mar 2019