Correcting for the individual patient regression to the mean effect.

Samuel Branders, PhD Guillaume Bernard, PhD Alvaro Pereira, PhD
Tools4Patient SA, Gosselies, Belgium

Introduction and Motivations

Often, the primary endpoint of RCTs is defined as a change from baseline of a continuous outcome. Such endpoint is subject to regression to the mean (RTM).

Regulators recommend including the outcome's baseline value as a covariate in the statistical analysis. However, this covariate cannot account completely for the RTM.

Using several similar baseline measures, we propose to model the RTM and to improve its correction.

Study Design and Patients

Two similar studies:

- Total 88 Peripheral Neuropathic Pain patients (30 + 58);
- 4 weeks of blinded placebo treatment b.i.d. as add-on therapy.

Pain evaluation:

- Average Pain Score (APS);
- Worst Pain Score (WPS);
- Severity sub-score of the Brief Pain Inventory (BPI-sev).
- Interference sub-score of the Brief Pain Inventory (BPI-interf).

The studies endpoints are define as the change from baseline of the pain evaluations (primary endpoint APS).

Regression to the mean

Regression to the mean (RTM) happens when repeated measures are observed in the same subjects.

Those measurements are observed with some variability around the true patient mean. In general, when the first observation of a subject is relatively high (or low), the second ones is likely to be less extreme and nearer the subject's true mean.

An endpoint defined as change from baseline (the difference between two repeated measures) is thus subject to the RTM.

Modelling the regression to the mean

To model the regression to the mean, we considered a situation with no treatment effect. The vectors of observed baseline values, x_1, and end-of-treatment values, x_2, can be modelled as follow :

$$x_1 = x_b + \epsilon_1, \quad \epsilon_1 \sim \mathcal{N}(0, \sigma_e^2) \tag{1}$$
$$x_2 = x_b + \epsilon_2, \quad \epsilon_2 \sim \mathcal{N}(0, \sigma_e^2) \tag{2}$$

Where x_b is a latent variable representing the true baseline value. The endpoint can then be expressed as follow :

$$y = x_2 - x_1 = \epsilon_2 - \epsilon_1 \tag{3}$$

Classicaly, the baseline x_1 is used as a covariate in the analyses. The correlation between the endpoint y and this covariate x_1 can be expressed as follow:

$$r(y, x_1) = \frac{Cov(y, x_1)}{\sigma_y \sigma_{x_1}} = \frac{\sigma_e^2}{\sqrt{2\sigma_e^2}\sqrt{\sigma_b^2 + \sigma_e^2}} \tag{4}$$
$$= \frac{1}{\sqrt{2}\sqrt{\frac{\sigma_b^2}{\sigma_e^2} + 1}} \tag{5}$$

In a linear model, the variance explained by x_1 is :

$$r^2(y, x_1) = \frac{1}{2\frac{\sigma_b^2}{\sigma_e^2} + 2} = \frac{1}{2SNR_b + 2} \tag{6}$$

Where $SNR_b = \frac{\sigma_b^2}{\sigma_e^2}$ is the signal-to-noise ratio of the x_1. The lower is the signal-to-noise ratio, the higher the explained variance will be.

Improving the regression to the mean correction

To increase the covariate correction, we have to decrease the signal-to-noise ratio of x_1. A new covariate, x_n, could computed by removing the signal x_b from the baseline x_1 :

$$x_n = x_1 - x_b \tag{7}$$
$$x_n = \epsilon_1 \tag{8}$$

We could estimate the hidden variable, x_b, while combined several similar baseline features. Those features are assumed to share similarities while having independent error terms.

Results

In our studies, we have 4 baseline measures (APS, WPS, BPI-Sev, and BPI-Interf) assessing the patient's baseline pain.

The latent variable x_b (true baseline APS) can be seen as a scaling of the patient's baseline pain. The 1st component of a PCA on the 4 baselines, x_{pca}, estimates this patient's pain :

$$\hat{x_b} = \beta x_{pca} + \alpha \tag{9}$$
$$\hat{x_n} = x_1 - \beta x_{pca} + \alpha \tag{10}$$

The new covariate, x_n, can be estimated as the residuals of the following linear model :

$$x_1 = \beta x_{pca} + \alpha + \epsilon \tag{11}$$
$$\hat{x_n} = \epsilon \tag{12}$$

We compared the variance explained on the primary endpoint (change from baseline of the APS).

Covariate	Adj r^2	p-value
x_1	10.8	0.001018
x_{pca}	-1.0	0.7526
x_n	28.5	5.35e-08

The new covariate x_n explains much more variance than the baseline value x_1.

Conclusion

In RCTs, the choice of primary endpoint is often

- conditioned by the recommendations of the agencies;
- subject to regression to the mean effect.

The gold standard is to include the baseline value of the endpoint as a covariate in the analysis. However, this covariate cannot account completely for the RTM.

Using several similar baseline covariates, we were able to improve the explained variance and the covariate correction.

This methodology is a first step improving the explained variance in RCTs. To go further, we should consider characterizing the treatment (or placebo) effect.

samuel.branders@tools4patient.com