Accounting for Regression to the Mean and Natural Growth in Uncontrolled Weight Loss Studies

William D. Johnson, Ph.D.
Pennington Biomedical Research Center

Regression to the Mean

- **Consider a study that enrolls children and adolescents ages 6 to 19 years in a study to investigate an intervention to reduce mean body mass index in kg/m^2 (BMI) during an observation period of, for example, 2 years.**
- **The focus of this presentation is on the analysis and interpretation of uncontrolled weight loss studies where a condition for enrollment is weight elevated to a level of being overweight or obese.**
- **In adults being overweight or obese is often defined as having BMI $\geq$ 30 regardless of age or gender; however, in children and adolescents overweight or obese is often defined as having BMI that exceeds the 85^{th} percentile for individuals in the U.S. with age and gender identical to the subject.**
- **Let Y_1 and Y_2 respectively denote BMI for a subject at the beginning (BOS) and end of study (EOS). The intervention administered during the study is assessed by comparing the mean BMI at EOS against the mean BMI at BOS.**
- **Even if there is no intervention effect, the mean BMI at EOS would be less than the mean BMI at BOS due to Regression to the Mean.**
- **Ideally participants would be randomized to the intervention or a comparative control but a surprising number of studies are uncontrolled.**

Regression to the Mean

1. Let Y_1 and Y_2 be bivariate normal random variables where

$$\mathrm{E}(Y_1) = \mu_1 \quad \text{and} \quad \mathrm{E}(Y_2) = \mu_2$$

$$\mathrm{Var}(Y_1) = \mathrm{Var}(Y_2) = \sigma^2 \text{ and } \mathrm{Cov}(Y_1, Y_2) = \rho\sigma^2$$

2. Consider the hypothesis

$$H_0\colon \mu_1 - \mu_2 = 0$$

3. The paired *t*-test would be the test statistic of choice.

Regression to the Mean

- **To derive analytic data that are absent variations attributable to age and gender, we use CDC Growth Charts to find each subject's gender and age-specific BMI percentile at BOS and at EOS (after their age increased at EOS).**
- **All percentiles are then converted to z-scores using the standard normal distribution [N(0,1)] to determine the *z*-scores for each specific subject's percentiles at BOS and EOS, respectively.**
- **Note that the BOS and EOS z-scores are two subsets of z-scores that would be distributed as N(0,1) only if the corresponding subsets of percentiles are representative samples of gender and age-specific percentiles for the U.S. population.**
- **Let X_1 and X_2 denote the resulting *z*-scores at BOS and EOS, respectively, and then standardize them to distributions with standard deviations =1 by the transformation $Z_1 = X_1/\delta$ and $Z_2 = X_2/\delta$ where δ is the common standard deviation for X_1 and X_2.**

Regression to the Mean

LA CaTS
LOUISIANA CLINICAL & TRANSLATIONAL SCIENCE CENTER

- **Suppose subjects are enrolled in the study only if their $Z_1 \geq z_c$ where $z_c \geq 0$ is called the BMI z-score cut-point for determining study eligibility.**
- **For example, suppose a requirement for enrollment in a study is $Z_1 \geq z_{85}$ where the cut-point z_{85}= 1.036 is the 85th percentile in the U.S. for a subject's gender and age.**
- **Thus, if the subject's BMI exceeds the 85th percentile he/she is defined to be overweight or obese and eligible for study provided other conditions are also satisfied.**

Regression to the Mean

LA CaTS
LOUISIANA CLINICAL & TRANSLATIONAL SCIENCE CENTER

Let θ denote the mean z-score at BOS, $\tau \leq 0$ the treatment effect and $\theta + \tau$ = the mean z-score at EOS. Further, let ρ_z denote the correlation of z-scores between BOS and EOS Thus,

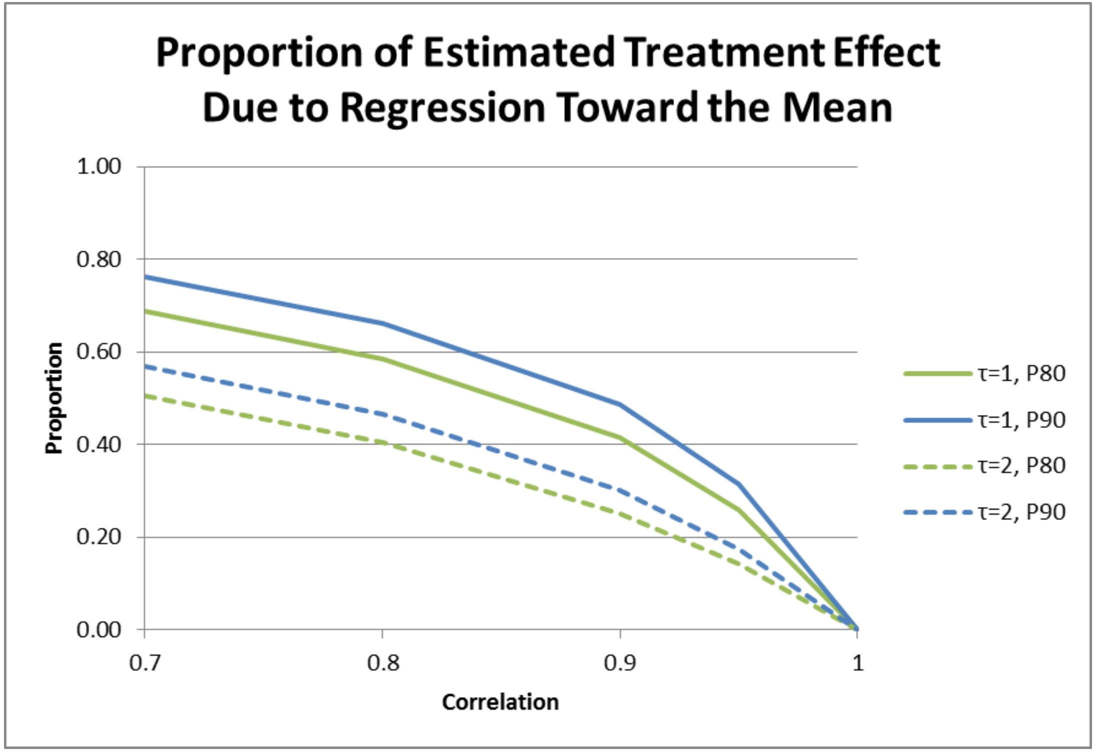
- $\mathrm{E}(Z_1) = \theta, \mathrm{E}(Z_2) = \theta + \tau,$
- $\mathrm{Var}(Z_1) = \mathrm{Var}(Z_2) = 1$ **and** $\mathrm{Cov}(Z_1, Z_2) = \rho_z$
- $\mathrm{E}(Z_1 | Z_1 \geq z_c) = \theta + \frac{q(z_c)}{Q(z_c)}$ **and** $\mathrm{E}(Z_2 | Z_1 \geq z_c) = \theta + \tau + \rho_z \frac{q(z_c)}{Q(z_c)}$
-
- $q(z_c) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-{z_c}^2}{2}\right)$ **and** $Q(z_c) = \int_{z_c}^{\infty} q(u)du$

Regression to the Mean

- $E(Z_2 - Z_1 \,|\, Z_1 \geq z_c) = \tau - (1 - \rho_z)\dfrac{q(z_c)}{Q(z_c)}$

- RTM $= -(1 - \rho_z)\dfrac{q(z_c)}{Q(z_c)}$ = regression to the mean
- Let $z_c = z_{0.85}$ = 1.036, q(1.036) =0.2332, Q(1.036) = 0.15, then for $\rho_Z = 1, 0.8, 0.5, 0.2, 0$:
- RTM = 0, 0.311, 0.777, 1.244, 1.555

Regression to the Mean



Concluding Remarks

LA CaTS
LOUISIANA CLINICAL & TRANSLATIONAL SCIENCE CENTER

- **Regression to the mean is an important consideration in the interpretation of intervention in weight loss studies where subjects are selected from the upper end of weight distribution.**
- **Ideally the study design would be randomized control trial.**
- **If good estimates of the parameters of the underlying un-truncated. distribution of pre-intervention weights (BMI) are available, good estimates of the effect weight loss intervention can be obtained even if a condition for study eligibility requires oe-intervention weight to be elevated.**
- **Normal growth during childhood can be accounted for by transforming weight (BMI) to gender and age-specific percentiles.**

Authors

William D. Johnson
Professor
Biostatistics
Pennington Biomedical Research Center
6400 Perkins Rd.
Baton Rouge, LA 70808
USA
Jeffrey.Burton@pbrc.edu

Jeff H. Burton
Assistant Professor
Biostatistics
Pennington Biomedical Research Center
6400 Perkins Rd.
Baton Rouge, LA 70808
USA
Jeffrey.Burton@pbrc.edu

Robbie Beyl
Assistant Professor
Biostatistics
Pennington Biomedical Research Center
6400 Perkins Rd.
Baton Rouge, LA 70808
USA
Robbie.Beyl@pbrc.edu

Supported by 1 U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health which funds the Louisiana Clinical and Translational Science Center