

Best (but oft-forgotten) practices: identifying and accounting for regression to the mean in nutrition and obesity research

Diana M Thomas,¹ Nicholas Clark,¹ Dusty Turner,¹ Cynthia Siu,² Tanya M Halliday,³ Bridget A Hannon,⁴ Chanaka N Kahathuduwa,⁵ Cynthia M Kroeger,^{6,7} Roger Zoh,⁸ and David B Allison⁷

¹Department of Mathematical Sciences, US Military Academy, West Point, NY, USA; ²Department of Data Science, COS and Associates Ltd., Hong Kong, China; ³Department of Health, Kinesiology, and Recreation, University of Utah, Salt Lake City, UT, USA; ⁴Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA; ⁵Department of Human Development and Family Studies, Texas Tech University, Lubbock, TX, USA; ⁶Charles Perkins Centre, School of Pharmacy, Faculty of Medicine and Health, University of Sydney, Sydney, Australia; ⁷Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, IN, USA; and ⁸School of Public Health, Indiana University, Bloomington, IN, USA

ABSTRACT

Background: Regression to the mean (RTM) is a statistical phenomenon where initial measurements of a variable in a nonrandom sample at the extreme ends of a distribution tend to be closer to the mean upon a second measurement. Unfortunately, failing to account for the effects of RTM can lead to incorrect conclusions on the observed mean difference between the 2 repeated measurements in a nonrandom sample that is preferentially selected for deviating from the population mean of the measured variable in a particular direction. Study designs that are susceptible to misattributing RTM as intervention effects have been prevalent in nutrition and obesity research. This field often conducts secondary analyses of existing intervention data or evaluates intervention effects in those most at risk (i.e., those with observations at the extreme ends of a distribution).

Objectives: To provide best practices to avoid unsubstantiated conclusions as a result of ignoring RTM in nutrition and obesity research.

Methods: We outlined best practices for identifying whether RTM is likely to be leading to biased inferences, using a flowchart that is available as a web-based app at <https://dustyturner.shinyapps.io/DecisionTreeMeanRegression/>. We also provided multiple methods to quantify the degree of RTM.

Results: Investigators can adjust analyses to include the RTM effect, thereby plausibly removing its biasing influence on estimating the true intervention effect.

Conclusions: The identification of RTM and implementation of proper statistical practices will help advance the field by improving scientific rigor and the accuracy of conclusions. This trial was registered at clinicaltrials.gov as NCT00427193. *Am J Clin Nutr* 2019;00:1–10.

Keywords: regression to the mean, statistical errors, nutrition and obesity research, treatment effect, unsupported conclusions

Introduction

Regression to the mean (RTM) is a statistical phenomenon that appears when repeated measurements of an outcome are taken (e.g., a pretest and a posttest) and when the outcome of interest is the change in the outcome of interest from pretest to posttest (i.e., posttest value – pretest value). RTM can make it appear that a treatment effect is present even in the absence of a treatment effect. The effects of RTM can lead to unsupported conclusions in studies, including studies of nutrition and obesity (1–3). Misattribution of RTM as an intervention effect typically, but not always, occurs when investigators restrict analyses in a segment of the sample above or below the population mean to determine pretest to posttest intervention changes (2, 4) and do not utilize a control group (5, 6). Unfortunately, if not accounted for, RTM can lead to incorrect conclusions.

This study was funded by the NIH (R25DK099080).

Supplemental Calculations and Supplemental Methods are available from the “Supplementary data” link in the online posting of the article and from the same link in the online table of contents at <https://academic.oup.com/ajcn/>. Data described in the manuscript, code book, and analytic code are made publicly and freely available without restriction in the Online Supporting Materials.

Address correspondence to DMT (e-mail: diana.thomas@westpoint.edu).

Abbreviations used: CALERIE 2, Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy Phase 2; NLSY, National Longitudinal Survey of Youth; RCT, randomized controlled trial; RTM, regression to the mean.

Received January 3, 2019. Accepted for publication July 23, 2019.

First published online 0, 2019; doi: <https://doi.org/10.1093/ajcn/nqz196>.

Unsupported conclusions in nutrition and obesity research

RTM was first recognized in 1886 by Sir Francis Galton (7). When evaluating parental and offspring heights, Galton observed that heights of offspring of taller parents were still greater than average, yet shorter than the heights of their parents (i.e., regressing towards the population mean). This phenomenon was initially described by Galton as “regression toward mediocrity” (7). Since the original reports on RTM by Sir Francis Galton, the effects of RTM have been detected, often postpublication, in numerous different research areas (8–11). RTM is encountered in studies of nutrition and obesity research for several reasons (Table 1).

First, obesity interventions are typically targeted to individuals who are classified with obesity (e.g., children above the 85th BMI percentile and adults with BMIs ≥ 30 kg/m²). Thus, by definition, obesity studies often restrict inclusion criteria for participants with a BMI or other cardiometabolic measures above the population mean (or at least exclude those from the low end of the distribution). Indeed, preoperative qualifications

for bariatric surgery generally require the patient to have a BMI >40 kg/m² (35 kg/m² with comorbidities), thereby introducing RTM effects in any analysis of bariatric surgery weight loss in an intervention study without a randomized control group.

A second reason may arise from a well-intentioned desire to ascertain a positive impact from an intervention that did not originally lead to a rejection of the null hypothesis. When this occurs, it is tempting to perform secondary analyses with those participants in the sample population who were most at risk. We draw out an example from (18) later on in this article. These exploratory analyses can provide interesting insights and help inform future studies and confirmatory hypothesis tests. Despite good intentions, when these subgroup analyses do not account for RTM, they can lead to misleading conclusions and misinformation about best practices for the management of diseases, such as obesity.

RTM also effects consumer perceptions, where individuals who are suffering from an illness or poor health may feel better after adhering to a fad diet, eliminating a food group, or restricting

TABLE 1 A sample of studies in nutrition and obesity research where ignoring the effect of regression to the mean led to unsupported conclusions¹

Study title	Description of RTM effect
Hypertension risk: exercise is medicine for most, but not all (12)	The study evaluated, through a secondary analysis of data, the individual differences that contributed to changes in blood pressure that resulted from exercise. After grouping data into low, moderate, and high changes in blood pressure, mean baseline measurements were examined between the 3 groups. The individuals with the largest decreases in blood pressure were found to have higher baseline blood pressure. This finding can be attributed to the effect of RTM (13).
RTM of repeated ambulatory blood pressure monitoring in 5 studies (14)	Five different studies that collected repeated blood pressure measurements were analyzed. Blood pressure readings that were high at baseline were lower during follow-ups and lower blood pressure readings at baseline were higher during follow-ups. The RTM effect was found for systolic and diastolic blood pressure and in both control and intervention groups.
Understanding the relationship between baseline BMI and subsequent weight changes in antipsychotic trials: effect modification or regression to the mean? (15)	The study tested the claim that antipsychotic agents led to less weight gain in individuals with high BMIs at baseline. A secondary analysis of the data revealed that the observed effect was not beyond what is expected due to RTM.
Strong, healthy, energized: striving for a healthy weight in an older, lesbian population (16)	The 12-wk intervention was designed for lesbian women 60 y of age and above with overweight/obesity. The study concluded the intervention was effective due to an increase in steps for the participants in the category with the lowest defined tertile of baseline step counts. The observed increase was likely due to RTM (1).
Uric acid lowering in relation to HbA1c reductions with the SGLT2 inhibitor tofogliflozin (17)	The study examined aggregated outcomes of 4 clinical trials on the effects of the SGLT2 inhibitor tofogliflozin (vs. placebo) on HbA1c and serum uric acid levels. The tofogliflozin effect was reported due to individuals with highest levels of HbA1c experiencing greater reductions in HbA1c than did those with lower baseline HbA1c levels. This effect was likely due to RTM (2).
Pre- and postevaluations of a weight management service for families with overweight and obese children, translated from the efficacious lifestyle intervention, PEACH (18)	The PEACH study was an adiposity-reducing lifestyle intervention in children aged 5–9 y. The study found statistically significant yet modest reductions in BMI and waist circumference z-scores, but did not include an intervention-free control group for comparison. When compared to longitudinal data from a study in children without an intervention (3), the PEACH study reductions in BMI z-scores were similar, suggesting that the observed changes were due to RTM (5).
Changes in telomere length 3–5 years after gastric bypass surgery (19)	This study found that telomere length increased in the lowest group of baseline telomere lengths postgastric bypass surgery. A statistically significant increase was not found in the entire sample. The study did not include a control group and the findings were attributable to a RTM effect rather than the intervention (20).
Do obese and extremely obese patients lose weight after lumbar spine fusions? Analysis of a cohort of 7303 patients from the Kaiser National Spine Registry (21)	The study concluded that patients with obesity were more likely to lose weight after spine fusion surgery, in comparison to those in lower BMI categories. Because there was no control group and subjects with the highest BMIs were isolated for the analysis, the conclusions can be explained by RTM (4).

¹ HbA1c, glycated hemoglobin; PEACH, Parenting, Eating and Activity for Child Health; RTM, regression to the mean; SGLT2, sodium glucose co-transporter 2.

TABLE 2 Summary of control participants¹

CALERIE 2 control group (<i>n</i> = 66)	Values
Age, y	37.88 ± 7.07
Height, cm	168.74 ± 8.27
First measurement baseline weight, kg	72.64 ± 8.68
Second measurement baseline weight, kg	72.29 ± 8.67
One-year weight, kg	72.38 ± 9.39
Two-year weight, kg	72.9 ± 9.16

¹Data are from the Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy Phase 2 (CALERIE 2). Data are presented as mean ± SD.

their diet to certain types of foods (22). While the consumer may attribute the improvement to changes in their diet, the improvement was most likely due to RTM.

What is regression to the mean?

Errors of inference due to RTM are often easily missed (8) and many find the phenomenon to be nonintuitive. Because theories are best understood with concrete examples, we illustrate RTM using data from the control group of the weight loss intervention, the Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy Phase 2 (CALERIE 2) (23). The CALERIE 2 study data are publicly available at <https://calerie.duke.edu/samples-data-access-and-analysis>.

The CALERIE 2 (23) study was designed to test the hypothesis that 2 y of caloric restriction at 25% below baseline levels would slow aging. The study was conducted at multiple sites: the Pennington Biomedical Research Center in Baton Rouge, LA; Tufts University in Boston, MA; and Washington University St. Louis, in Missouri. A sample of 220 participants was randomized into a caloric restriction group or a control group. The participants in the control group were advised to continue their routine daily diets. Body weights were measured 4 times during the study: twice at baseline, at 1 y, and at 2 y. Our analysis was restricted to the participants in the control group that had a full set of weight measurements at each time point. A summary of participant characteristics appears in Table 2. From Table 2, we see that the mean weights and SDs at each time point in the control sample of CALERIE 2 remained stable, and a series of Student's *t*-tests with Bonferroni corrections found no significant differences in mean body weights among time points.

We would not expect the measured weights of an individual control participant in CALERIE 2 to be exactly the same at each time point (24); however, over time, the repeated measured weights of an individual would yield their own distribution, with a mean and an SD (25). If we selected a series of subjects whose initial weights were among the highest of all the original weights, then the second weight for each individual would very likely be lower in magnitude than the first and closer to the mean of the individual's repeated measures. We say that the second weight, then, "regressed to the mean."

Figure 1 was developed by dividing the first baseline measured weights in the control sample of CALERIE 2 into thirds. The upper third (blue in Figure 1) represents individuals whose measured weights were 1 SD higher than the mean in the first baseline measurement. Likewise, the lower third (red in Figure 1) represents individuals with measured weights 1 SD below

the mean. Finally, the middle third (green in Figure 1) are individuals with measured weights within 1 SD of the mean. Upon repeated measurements, we can see the blue and red marked individuals "move" toward the mean. The key point is that this movement was not due to the intervention effect. Unsupported conclusions because of failing to account for RTM occur when the observed mean change between these repeated measurements in a sample above or below the population mean is interpreted as an intervention effect.

Next, we outlined methods to determine the existence of RTM and how to quantify the degree of RTM effect. We also include study design considerations that account for the effects of RTM.

Methods

Determining whether an analysis is susceptible to regression to the mean

After identifying the presence of RTM, leading to unsupported conclusions in numerous studies (1, 2, 4, 6, 15, 20), we developed a flowchart for investigators, editors, and referees to determine whether RTM effects may be biasing estimates of intervention effects (Figure 2A). A web-based application, available at <https://dustyturner.shinyapps.io/DecisionTreeMeanRegression/>, was designed to automate the flowchart, specific to the user's needs. We demonstrated how to apply the flowchart in Figure 2A, using a specific example.

Example.

A recent study in middle-school children in a large, urban area evaluated the effectiveness of a physical activity intervention, the We Run This City Youth Marathon Program (26). The program's purpose was to improve markers of health by increasing physical activity through structured running or walking for 12–14 wk. Specific markers, measured at pretest and posttest, were BMI, waist-to-hip ratio, blood pressure, and fitness levels. The authors found no significant mean change in BMI percentiles across the entire sample. However, when the sample was stratified by preintervention BMI percentiles, and separate analyses were conducted within each strata, a significant change between pretest to posttest BMIs was identified in the subgroup of participants classified with overweight or obesity. This is a classic example of how RTM is misattributed to an intervention effect.

Determining whether RTM is a problem in this context can be done by responding to the questions in our RTM decision tree. The original intervention (26) did not restrict their sample based on any criteria. Thus, the response to the first flowchart question, "were subjects receiving the intervention being studied or sampled in such a way that those on 1 side of the mean were overrepresented (e.g., only including subjects with BMIs above some level)?," is "no," moving us to the left of the flowchart. However, the answer to the next question, "are analyses being conducted or conclusions being drawn on subsets of the intervention group defined by baseline distribution of the outcome variables?," is "yes," because the secondary analysis was restricted solely to participants on the upper end of the baseline BMI percentiles. As a control group was not included, the answer to the next question is

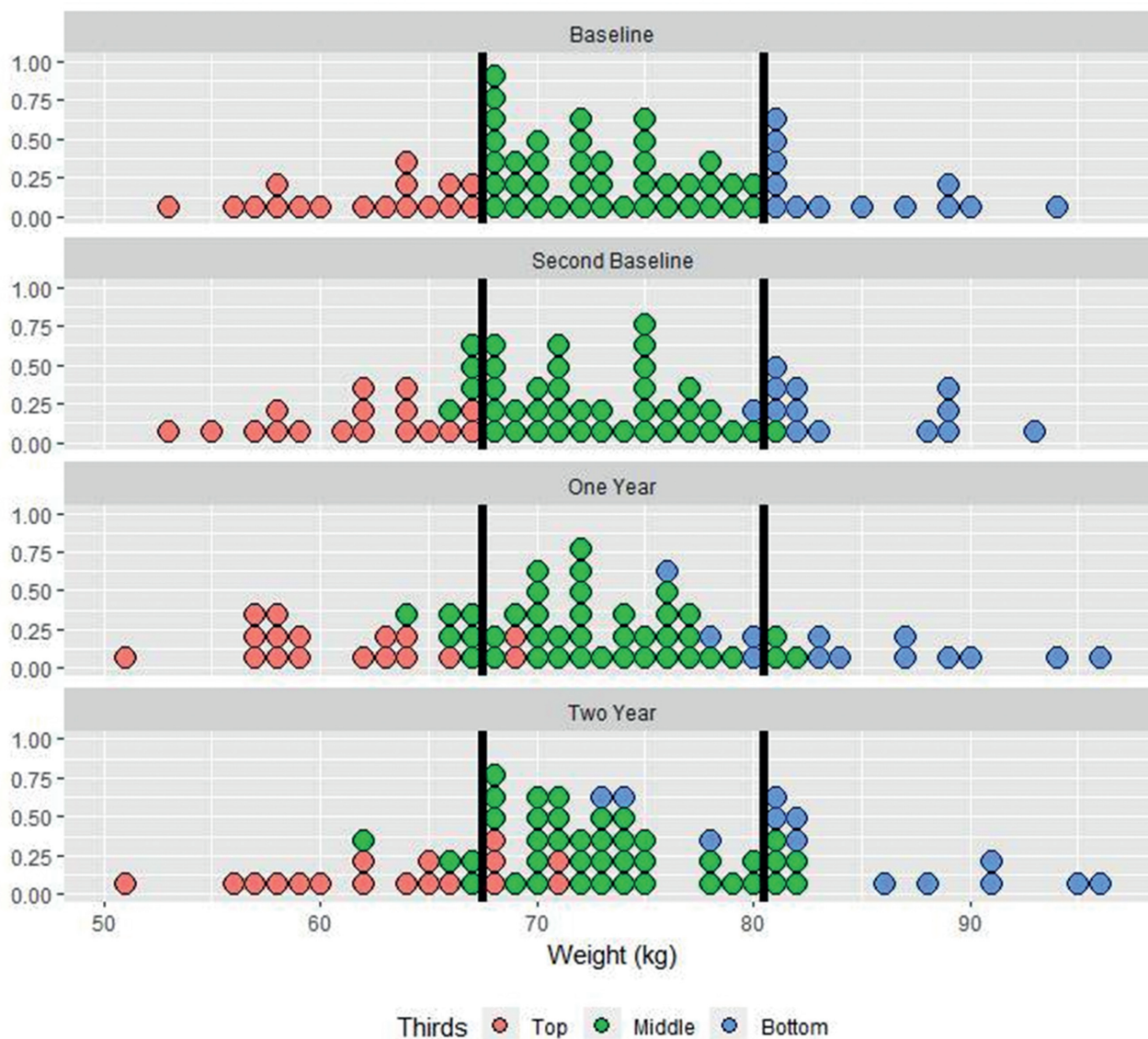


FIGURE 1 Distribution of control population weights in the Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy Phase 2 (CALERIE 2) study at baseline, with second measurements at baseline, 1 y, and 2 y. The control participants in the first baseline measurement were color coded by location: blue for 1 SD above the mean, red for 1 SD below the mean, and green for within 1 SD of the mean. The location of participants in subsequent measurements can be tracked. We found that the blue-coded participants moved down toward the mean, and the red-coded participants moved up toward the mean, visually demonstrating the effect of regression to the mean.

“no.” Because the conclusions were being drawn by testing for any significance in changes in BMIs from pretest to posttest in the subset of participants with a BMI that was classified as overweight or obese, the answer to the next question is “yes.” This leads to the final outcome in the followed path (Figure 2B), “RTM is a problem and could be creating spurious conclusions.”

Within-subject variance using repeated measures

Different yet similar approaches to calculate the RTM effect have appeared in the literature (15, 25, 27–29). A derivation to calculate the effect of RTM, assuming the data are normally distributed, appeared in a work by Davis (29), and we refer

the reader interested in full mathematical details of this specific formula to the article.

The RTM effect when the 2 variables in question are assumed to have the same mean and variance are computed using the following formula (25, 29):

$$\text{RTM effect} = \frac{\sigma_w^2}{\sqrt{\sigma_w^2 + \sigma_b^2}} G(z) = \frac{\sigma_w^2}{\sqrt{\sigma_t^2}} G(z) \quad (1)$$

where σ_w^2 is the within-subject variance, σ_b^2 is the between-subject variance, and σ_t^2 is the total variance. The function, $G(z)$,

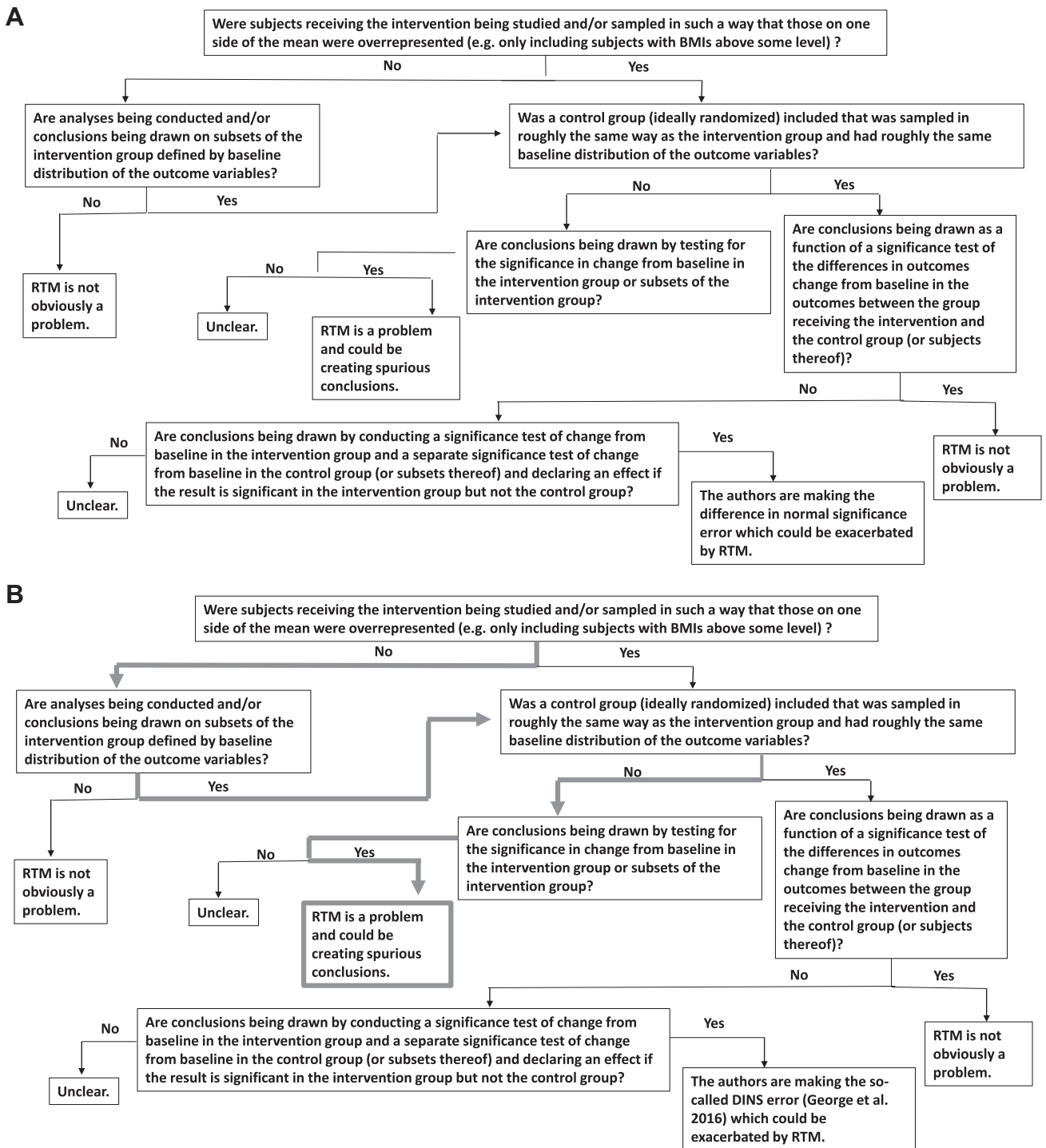


FIGURE 2 (A) Decision tree for assessing whether RTM leads to unsubstantiated conclusions above intervention effects. (B) The path is highlighted in the decision tree that follows the analysis performed in reference 26. RTM, regression to the mean.

is defined as:

$$G(z) = \frac{\phi(z)}{1 - \Phi(z)} \quad (2)$$

where $\phi(z)$ and $\Phi(z)$ are the standard normal probability density function and cumulative distribution functions, respectively, and $z = \frac{|c - \mu|}{\sigma_w^2 + \sigma_b^2}$, where c represents the cut-off threshold in the data

of the baseline measurement that is of interest and μ is the mean of the baseline measurements. For example, if we were interested in only the participants in an adult data set with obesity, then, c would equal 30 kg/m².

Equation (1) is worth examining to see how the terms in the product impact the RTM effect. For example, we expect that the RTM effect should be larger if there is larger intraindividual variability: that is, greater within-subject variance. This is the

same as saying the RTM effect increases when the correlation between the repeated measurements is low. We also would expect that the RTM effect should be larger with an increasing threshold value of c : that is, as the subjects being studied have progressively more extreme values.

To gain these insights, we would begin with the function $G(z)$, which is increasing in z . This implies that the further c is in distance away from the mean, the larger in value $G(z)$ and, consequentially, the larger the RTM effect. More simply, partitioning the data for analysis further away from the mean will increase the RTM effect.

If we divide the numerator and denominator of the first term in Equation (1) by σ_w^2 , this term reformulates as:

$$\frac{\sigma_w^2}{\sqrt{\sigma_w^2 + \sigma_b^2}} = \frac{1}{\sqrt{\frac{1}{\sigma_w^2} + \frac{\sigma_b^2}{\sigma_w^4}}} \quad (3)$$

Considering σ_b^2 to be fixed, we can see in the above formulation that a large σ_w^2 will result in a denominator close to 0, which will yield a large first term in the product in Equation (1). So a larger within-subject variance will increase the RTM effect.

Finally, dividing the numerator and denominator of the first term in Equation (1)'s product by σ_b^4 reformulates the term as:

$$\frac{\sigma_w^2}{\sqrt{\sigma_w^2 + \sigma_b^2}} = \frac{\left(\frac{\sigma_w}{\sigma_b}\right)^2}{\sqrt{\left(\frac{\sigma_w}{\sigma_b}\right)^2 + 1}} \quad (4)$$

From here, we see that as σ_b gets larger relative to σ_w (i.e., if the intersubject variability is much larger than the intrasubject variability), RTM becomes lower. Thus, RTM is directly proportional to the intrasubject variability, and inversely proportional to the ratio between the intra- and intersubject variabilities.

We illustrated an estimation of the RTM effect using the CALERIE 2 control group data set, composed of 66 subjects who had their weight measured 4 times: twice at baseline, once at 12 mo, and once at 24 mo. No data were missing; all calculations are available in the Excel file supplied in the **Supplemental Calculations**. The supplied Excel sheet contains the Excel formulas for each calculation, which can be viewed by selecting the cell.

Example.

We first needed to set a cut-point which we designated as $c = 81.31$, which was 1 SD above the mean (blue dots in Figure 1).

To illustrate how to estimate the RTM effect using the formula (1), we considered the "groups" as the individual subjects in CALERIE 2. For $\hat{\sigma}_w^2$, we first computed the degrees of freedom within groups, which equaled the total number of observations less the number of groups: $264 - 66 = 198$. Next, we computed the sum of squared differences between each group's mean and the individual observations within the group. To compute the within-subject variance, we added the 66 squared differences, and divided the sum by the within-group degrees of freedom, 242, yielding $\hat{\sigma}_w^2 = 5.76$.

To estimate $\hat{\sigma}_b^2$, we first needed the between-group degrees of freedom, which equaled the number of groups (66 individuals) minus 1, or $df: 66 - 1 = 65$. We next calculated the between-group mean square error by summing the squared difference between each group mean and the overall grand mean, 72.55. This sum was multiplied by the sample size in each group, which was 4, and divided by the degrees of freedom to arrive at the between-group mean square error (see Supplemental Calculations for expanded details):

$$MSB = 305.25 \quad (5)$$

Finally, substituting the values of MSB and $\hat{\sigma}_w^2$ resulted in:

$$\hat{\sigma}_b^2 = \left(\frac{1}{4}\right)(MSB - \hat{\sigma}_w^2) = 74.87 \quad (6)$$

Next, we calculated the ratio:

$$G(z) = \frac{\phi(z)}{1 - \Phi(z)} \quad (7)$$

The value of z was determined by substituting the cut-point we had designated at the beginning of the example, $c = 81.31$, and $\hat{\sigma}_w^2$ and $\hat{\sigma}_b^2$ into $z = \frac{|c - \bar{X}|}{\sqrt{\hat{\sigma}_w^2 + \hat{\sigma}_b^2}}$, to arrive at $z = 0.10$.

So,

$$G(0.10) = \frac{\phi(0.10)}{1 - \Phi(0.10)} = 0.87 \quad (8)$$

Thus, the estimated RTM effect is:

$$\left(\frac{5.76}{\sqrt{5.76 + 74.87}}\right) 0.87 = 0.56 \text{ kg} \quad (9)$$

To calculate the RTM-free effect (the part of the observed change score that may have been due to the intervention and was uninfluenced by RTM) in the intervention group, restricted to those individuals above a baseline weight of 81.31 kg ($n = 36$), we took the total magnitude of weight loss in this group, 6.28 kg, and subtracted the estimated RTM effect: $6.28 - 0.56 = 5.72$ kg (25).

In the case of nonnormal distributions, derivations to account for the RTM effect also exist (30).

Using nationally representative longitudinal population data

A recent commentary on RTM (6) used national longitudinal survey data to determine whether the magnitude of the change in BMI z-scores from pretest to posttest in a study by Burke et al. (31) was less than would be expected if the change were due to RTM alone. Burke et al. (31) evaluated changes in BMI z-scores in children participating in a school-based program designed to improve healthy behaviors. In a recent commentary on RTM, Skinner et al. (6) used the 1997 cohort of the National Longitudinal Survey of Youth (NLSY), which is a longitudinal, nationally representative group of children, measured in 2-y intervals. Skinner et al. (6) looked at the 1997 cohort (<http://www.bls.gov/nls/nlsy97.htm>) and measured height and weight values

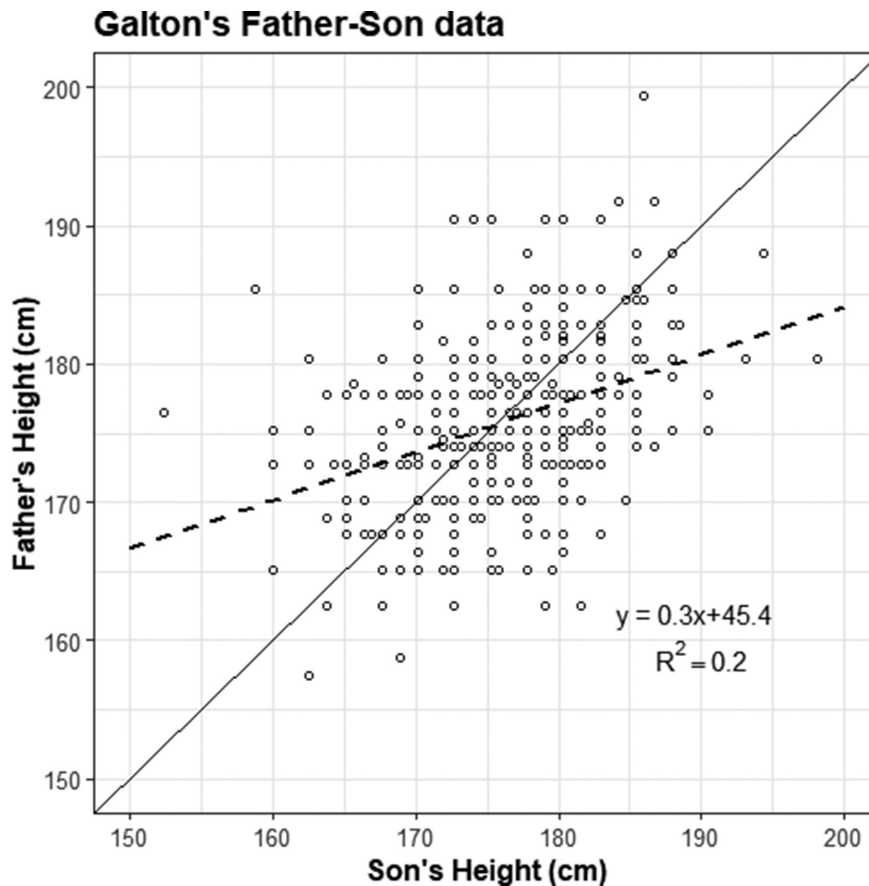


FIGURE 3 The graph shows the fathers' heights (cm) on the ordinate and the sons' heights on the abscissa from Galton's study (7). The tallest fathers had shorter sons and, likewise, the shortest sons had taller fathers. The deviation of the line of regression from the line of identity reflects the degree of regression to the mean (RTM).

at age 9 and age 11 to approximate the time duration of the Burke et al. (31) study.

Skinner et al. (6) found that in the NLSY data set, girls with obesity had an overall decrease of -0.22 in the BMI z-score. For boys with obesity, there was a -0.21 decline in the BMI z-score. In the school-based intervention, Burke et al. (31) found that girls with obesity's BMI z-scores were reduced by -0.10 . Boys demonstrated a decline of -0.12 . Because these were both within the -0.22 for girls and -0.21 for boys found in NLSY, the observed changes in Burke et al. (31) did not go beyond those accounted for by RTM. Thus, the changes observed by Burke et al. (31) did not support an intervention effect.

Using graphs

We can visually observe the effects of RTM by plotting pairs of first and second measurements with the first measurement on the ordinate and the second measurement on the abscissa. If there were no difference between the first and second measurements, the pairs would fall along the line of identity. In the presence of RTM, the line of regression would skew away from the line of identity. One of the best examples of this can be drawn from Galton's (7) observations in father-son height data. A reproduction of the plot of fathers' heights versus sons' heights, obtained from a publicly available database (www.randomservice

[es.org/random/](http://www.randomservice.org/random/)), appears in Figure 3. Figure 3 demonstrates how the difference between the line of regression and line of identity increases as fathers' heights become taller or shorter compared to the median value. The distance between the lines represents a quantification of the RTM effect as a continuous function of the fathers' heights.

Statistical tests related to testing for effects

Several statistical tests or procedures have been offered that are germane to RTM. These tests evaluate different properties of baseline and posttest data. Some test for the presence of or estimate the magnitude of some aspect of RTM. Some test for the differences in the mean or variance between the distributions of 2 different variables (e.g., a baseline measurement and an outcome measurement, such as BMI) beyond what can be attributable to RTM. However, we note the following. There can be cases in which 1) there is a change in the mean, variance, or other aspect of the distribution of a variable after a treatment or intervention is applied; and 2) the change cannot be attributed solely to RTM; yet 1) and 2) both being true does not necessarily imply that the treatment or intervention has had an effect. This is because there can be factors other than RTM and treatment effects which produce such changes in variables' distributions (32). Finally, there are other tests which

TABLE 3 Tests germane to regression to the mean¹

Test	Alternative hypothesis	Test	Probative for causal effects of treatment?
Baseline-change association (33)	There is an association of baseline values of the variable (e.g., BMI) with changes in the variable after treatment.	Ordinary tests of correlation (e.g., Pearson's product-moment correlation) and association between baseline values and changes.	No
Variance reduction (27, 34)	The distribution of the variable has, on average, smaller deviations from the mean at the second time point than it did at the first; that is, there is a trend toward "mediocrity."	Tests of reduction in variance.	No
Excess mean change (28)	There is a decrease [increase] in the mean of a variable among subjects selected to be (on average) above [below] the mean of that variable that cannot be explained solely by RTM under the assumption of a bivariate normal distribution.	Tests a parameter, denoted γ , which quantifies the change in the mean subsequent to treatment and after accounting for the expected change in the outcome due solely to RTM.	No
Control-adjusted mean change (35, 36)	An intervention had an effect on a change in an outcome variable.	Standard comparison of mean changes in outcome between treatment and control groups, with treatment being assigned at random (35).	Yes ²
Treatment assignment \times baseline interaction (15, 37)	An intervention had different effects on change in a variable as a function of the baseline values of that variable.	Tests of interaction between baseline values and treatment, with changes as outcomes.	Yes ²

¹Table does not contain an exhaustive listing. Data are for situations involving the study of changes in an outcome variable after a treatment is applied between baseline and endpoint measurement. RTM, regression to the mean.

²If assignment was randomized.

determine whether there is an effect of treatment on an outcome and whether the effect of treatment depends on the baseline level of the outcome variable. We summarized selected tests in **Table 3**.

The Baseline-Change Association test simply entails ordinary tests of a correlation (e.g., Pearson's product-moment correlation) or association between baseline values and changes. Such tests are merely tests of an association and provide no evidence about whether treatment effects depend on baseline values or even whether treatment effects exist.

The Variance Reduction test can be utilized to assess whether there is a "pulling in" of the extremes of the distribution toward population norms (38). RTM occurs when the absolute value of the correlation between the baseline and endpoint (ρ) is less than 1.0: that is, $|\rho| < 1.0$. Thus, stating that RTM exists is equivalent to stating that $|\rho| < 1.0$. Notably, the condition that $|\rho| < 1.0$: 1) does not imply a reduction in variance (i.e., that the variance is smaller) between the baseline and endpoint; and 2) is compatible with there being a reduction in variance between the baseline and endpoint. In other words, RTM can exist either with or without the presence of a variance reduction. Hence, while $0 \leq \rho < 1.0$ implies that, on average, those with the highest initial values will have their scores go down the most (or up the least) and those with the lowest initial values will have their scores go up the most (or down the least) (33), this does not imply a less dispersed distribution (39), where by less dispersed we mean that there are, on average, smaller squared deviations from the center of the distribution of the outcome variable after treatment. If we wish to assert that there is less dispersed distribution of the outcome after treatment, then we can test for reductions

in variance. Testing for a reduction in dispersion is of interest in multiple domains. For example, those studying the potential effects of social norms on behaviors (38) and the potential effects of breastfeeding on offspring BMI (40) had posited that the interventions might both increase the lower percentiles of an outcome distribution while decreasing the upper percentiles. If we wish to assert that there is such a reduction in the dispersion of the outcome after treatment, then we can test for reductions in variance. Note that any reduction in variance after a treatment is applied could be due to many factors and does not, in and of itself, imply any effect of the treatment on producing this reduction in dispersion.

The Excess Mean Change, or γ , test evaluates a parameter, denoted γ , which quantifies the change in the mean after treatment after accounting for the expected change in the outcome due solely to RTM in samples selected for deviating from the mean at baseline. The test is more heavily dependent on assumptions than some of the other tests described here. In addition, it only tests for whether there is a change in the mean beyond what can be accounted for by RTM. It does not demonstrate that any such changes are due to effects of treatment.

Finally, the Control-Adjusted Mean Change and Treatment Assignment \times Baseline Interaction tests are just the "good, old-fashioned" tests generally applied to controlled trials, and they test for treatment effects and differential treatment effects as a function of baseline values, respectively. Notably, these tests require comparing control to treatment groups, for causal inference with treatment assignment being randomized. Historic nonrandomized controls can be used as comparators, but then causal inferences are not strictly justified (6).

Results

The gold standard to account for RTM and avoid numerous other statistical errors that do not support conclusions (37, 41–43) is to design interventions that are randomized and include a control group. There are some cases where it may not be feasible or possible to design a randomized controlled trial (RCT). Perhaps the expense of a trial is too exorbitant or the setting may present an uncontrolled experiment (44) in which it is not feasible or ethical to conduct an RCT. In these cases, the RTM effect can either be estimated from repeated measurements at baseline or through estimates from comparable national databases or other existing data sets, such as those performed in other studies (6, 43). In a case where there are >2 repeated baseline measurements, the RTM effect can be derived to include all repeated measurements. This derivation is included in the **Supplemental Methods**.

A second case where it is not possible to evaluate data from a control group is in a secondary analysis of existing data from non-RCT interventions. In these cases, investigators first need to determine whether their analysis may include an RTM effect. We suggest using the decision tree provided here (Figure 2) and available as a web-based app (<https://dustyturner.shinyapps.io/DecisionTreeMeanRegression/>) to first determine whether the analysis may lead to conclusions impacted by RTM. If RTM is thought to be present, quantitative estimates similar to those made here may be performed to identify whether the apparent intervention effect is plausibly accountable for solely by RTM.

Discussion

The error of overlooking RTM as a possible etiology of what is mistakenly believed to be an intervention effect is not unique to nutrition and obesity research (8, 9, 11). Examples of RTM causing bias in the analyses of sports performance and other data have been highlighted in Kahneman's *Thinking, Fast and Slow* (45). As Kahneman pointed out (45), when observing the effects of RTM, a false narrative is often developed around the observed effect. For example, Secrist, a statistician, discovered that businesses that did really well financially in 1 y had less stellar performance the next year (46). Likewise, he found that poorly performing businesses improved the following year. From his observation, Secrist concluded that the trend would continue and all businesses would be performing at a mediocre level, hence the title of his book, *The Triumph of Mediocrity in Business* (46). Another more recent example appeared in a study on whether social norm messaging could induce household energy conservation. The messaging provided households with their previous week's energy consumption, along with the average neighborhood energy consumption. From follow-up measurements, it was observed that households who consumed higher than average energy at baseline decreased consumption and households who consumed lower than average energy at baseline increased consumption (10). Because there was no control group, it was not possible to definitively conclude whether there was an intervention effect beyond RTM (38).

In conclusion, we have presented strategies for defining, identifying, and correcting RTM in nutrition and obesity research. This phenomenon continues to be an issue in the fields of nutrition and obesity research, as ignoring RTM can lead to spurious conclusions and cause researchers to exaggerate intervention

effects or create them where none exist. We hope that through the examples and tools presented herein, researchers will be more aware of RTM and consider it in the designs and analyses of future work.

The authors' responsibilities were as follows—DMT, DBA: designed the research and had primary responsibility for the final content; DT: developed the web-based app; DMT, NC, RZ: analyzed the data and performed the statistical analysis; DMT, NC, DBA: wrote the paper; all authors: reviewed and analyzed the literature; and all authors: read and approved the final manuscript. The authors had no conflicts of interest.

References

- Halliday TM, Thomas DM, Siu CO, Allison DB. Letter to the editor. *J Women Aging* 2018;30(1):2–5.
- Kahathuduwa CN, Thomas DM, Siu C, Allison DB. Unaccounted for regression to the mean renders conclusion of article titled “Uric acid lowering in relation to HbA1c reductions with the SGLT2 inhibitor tofogliflozin” unsubstantiated. *Diabetes Obes Metab* 2018;20(8):2039–40.
- Cockrell Skinner A, Goldsby TU, Allison DB. Regression to the mean: A commonly overlooked and misunderstood factor leading to unjustified conclusions in pediatric obesity research. *Child Obes* 2016;12(2):155–8.
- Kroeger CM, Allison DB, Thomas DM, Siu CO. To the editor. *Spine (Phila Pa 1976)* 2018;43(8):E492–E3.
- Hannon BA, Thomas DM, Siu C, Allison DB. The claim that effectiveness has been demonstrated in the parenting, eating and activity for child health (PEACH) childhood obesity intervention is unsubstantiated by the data. *Br J Nutr* 2018;120(8):958–9.
- Skinner AC, Heymsfield SB, Pietrobelli A, Faith MS, Allison DB. Ignoring regression to the mean leads to unsupported conclusion about obesity. *Int J Behav Nutr Phys Act* 2015;12:56.
- Galton F. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland* 1886;15:246–63.
- Mokros A, Habermeyer E. Regression to the mean mimicking changes in sexual arousal to child stimuli in pedophiles. *Arch Sex Behav* 2016;45(7):1863–7.
- Lee J, Chung K, Kang S. Evaluating and addressing the effects of regression to the mean phenomenon in estimating collision frequencies on urban high collision concentration locations. *Accid Anal Prev* 2016;97:49–56.
- Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. *Psychol Sci* 2007;18(5):429–34.
- Jones HE, Spiegelhalter DJ. Accounting for regression-to-the-mean in tests for recent changes in institutional performance: Analysis and power. *Stat Med* 2009;28(12):1645–67.
- Loenneke JP, Fahs CA, Abe T, Rossow LM, Ozaki H, Pujol TJ, Bembien MG. Hypertension risk: Exercise is medicine* for most but not all. *Clin Physiol Funct Imaging* 2014;34(1):77–81.
- Atkinson G, Loenneke JP, Fahs CA, Abe T, Rossow LM. Individual differences in the exercise-mediated blood pressure response: Regression to the mean in disguise? *Clin Physiol Funct Imaging* 2015;35(6):490–1.
- Moore MN, Atkins ER, Salam A, Callisaya ML, Hare JL, Marwick TH, Nelson MR, Wright L, Sharmar JE, Rodgers A. Regression to the mean of repeated ambulatory blood pressure monitoring in five studies. *J Hypertens* 2019;37(1):24–9.
- Allison DB, Loebel AD, Lombardo I, Romano SJ, Siu CO. Understanding the relationship between baseline BMI and subsequent weight change in antipsychotic trials: Effect modification or regression to the mean? *Psychiatry Res* 2009;170(2–3):172–6.
- Tomisek A, Flinn B, Balsky T, Gruman C, Rizer AM. Strong, healthy, energized: Striving for a healthy weight in an older lesbian population. *J Women Aging* 2017;29(3):230–42.
- Ouchi M, Oba K, Kaku K, Suganami H, Yoshida A, Fukunaka Y, Jutabha P, Morita A, Otani N, Hayashi K, et al. Uric acid lowering in relation to HbA1c reductions with the SGLT2 inhibitor tofogliflozin. *Diabetes Obes Metab* 2018;20(4):1061–5.

18. Moores CJ, Miller J, Daniels LA, Vidgen HA, Magarey AM. Pre-post evaluation of a weight management service for families with overweight and obese children, translated from the efficacious lifestyle intervention parenting, eating and activity for child health (PEACH). *Br J Nutr* 2018;119(12):1434–45.
19. Dershem R, Chu X, Wood GC, Benotti P, Still CD, Rolston DD. Changes in telomere length 3–5 years after gastric bypass surgery. *Int J Obes (Lond)* 2017;41(11):1718–20.
20. Smith DL Jr., Thomas DM, Siu CO, Verhulst S, Allison DB. Regression to the mean, apparent data errors and biologically extraordinary results: Letter regarding “changes in telomere length 3–5 years after gastric bypass surgery.” *Int J Obes (Lond)* 2018;42(4):949–50.
21. Akins PT, Inacio MC, Bernbeck JA, Harris J, Chen YX, Prentice HA, Guppy KH. Do obese and extremely obese patients lose weight after lumbar spine fusions? Analysis of a cohort of 7303 patients from the Kaiser National Spine Registry. *Spine (Phila Pa 1976)* 2018;43(1):22–7.
22. Warner A. *The Angry Chef’s Guide to Spotting Bullsh*t in the World of Food: Bad Science and the Truth About Healthy Eating: The Experiment*. New York, NY: The Experiment, LLC; 2018.
23. Rickman AD, Williamson DA, Martin CK, Gilhooly CH, Stein RI, Bales CW, Roberts S, Das SK. The CALERIE study: Design and methods of an innovative 25% caloric restriction intervention. *Contemp Clin Trials* 2011;32(6):874–81.
24. Heymsfield SB, Peterson CM, Thomas DM, Hirezi M, Zhang B, Smith S, Bray G, Redman L. Establishing energy requirements for body weight maintenance: Validation of an intake-balance method. *BMC Res Notes* 2017;10(1):220.
25. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: What it is and how to deal with it. *Int J Epidemiol* 2005;34(1):215–20.
26. Borawski EA, Jones SD, Yoder LD, Taylor T, Clint BA, Goodwin MA, Trapl ES. We run this city: Impact of a community-school fitness program on obesity, health, and fitness. *Prev Chronic Dis* 2018;15:E52.
27. Kelly C, Price TD. Correcting for regression to the mean in behavior and ecology. *Am Nat* 2005;166(6):700–7.
28. Senn SJ, Brown RA. Estimating treatment effects in clinical trials subject to regression to the mean. *Biometrics* 1985;41(2):555–60.
29. Davis CE. The effect of regression to the mean in epidemiologic and clinical studies. *Am J Epidemiol* 1976;104(5):493–8.
30. Schwarz W, Reike D. Regression away from the mean: Theory and examples. *Br J Math Stat Psychol* 2018;71(1):186–203.
31. Burke RM, Meyer A, Kay C, Allensworth D, Gazmararian JA. A holistic school-based intervention for improving health-related knowledge, body composition, and fitness in elementary school students: An evaluation of the HealthMPowers program. *Int J Behav Nutr Phys Act* 2014;11:78.
32. Campbell DT, Stanley JC, Gage NL. *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally; 1966.
33. Clifton L, Clifton DA. The correlation between baseline score and post-intervention score, and its implications for statistical analysis. *Trials* 2019;20(1):43.
34. Hotelling H. Review: The triumph of mediocrity in business. *J Am Statist Assoc* 1933;28(184):463–5.
35. Bland JM, Altman DG. Best (but oft forgotten) practices: Testing for treatment effects in randomized trials by separate analyses of changes from baseline in each group is a misleading approach. *Am J Clin Nutr* 2015;102(5):991–4.
36. O’Connell NS, Dai L, Jiang Y, Speiser JL, Ward R, Wei W, Carroll R, Gebregziabher M. Methods for analysis of pre-post data in clinical research: A comparison of five common methods. *J Biom Biostat* 2017;8(1):1–8.
37. George BJ, Brown AW, Allison DB. Errors in statistical analysis and questionable randomization lead to unreliable conclusions. *J Paramed Sci* 2015;6(3):153–4.
38. Verkooijen KT, Stok FM, Mollen S. The power of regression to the mean: A social norm study revisited. *Eur J Soc Psychol* 2015;45:417–25.
39. Stigler S. Darwin, Galton and the statistical enlightenment. *J R Stat Soc A* 2010;173(3):469–82.
40. Beyerlein A, Toschke AM, von Kries R. Breastfeeding and childhood obesity: Shift of the entire BMI distribution or only the upper parts? *Obesity (Silver Spring)* 2008;16(12):2730–3.
41. Allison DB, Brown AW, George BJ, Kaiser KA. Reproducibility: A tragedy of errors. *Nature* 2016;530(7588):27–9.
42. George BJ, Beasley TM, Brown AW, Dawson J, Dimova R, Divers J, Goldsby TU, Heo M, Kaiser KA, Keith SW, et al. Common scientific and statistical errors in obesity research. *Obesity (Silver Spring)* 2016;24(4):781–90.
43. Senn S. Statistical pitfalls of personalized medicine. *Nature* 2018;563(7733):619–21.
44. Hoyt RW, Friedl KE. Field studies of exercise and food deprivation. *Curr Opin Clin Nutr Metab Care* 2006;9(6):685–90.
45. Kahneman D. *Thinking, Fast and Slow*. 1st ed. New York, NY: Farrar, Straus and Giroux; 2013.
46. Secrist H. *The Triumph of Mediocrity in Business*. Evanston, IL: Bureau of Business Research, Northwestern University; 1933.