Neglecting regression to the mean continues to lead to unwarranted conclusions: Letter regarding “The magnitude of weight loss induced by metformin is independently associated with BMI at baseline in newly diagnosed type 2 diabetes: Post-hoc analysis from data of a phase IV open-labeled trial”\textsuperscript{1}

Bridget A. Hannon\textsuperscript{1,A,D–F}, Diana M. Thomas\textsuperscript{2,D–F}, Cynthia O. Siu\textsuperscript{3,E,F}, David B. Allison\textsuperscript{4,A,D–F}

\textsuperscript{1} Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, USA
\textsuperscript{2} Department of Mathematical Sciences, United States Military Academy, West Point, USA
\textsuperscript{3} COS & Associates Ltd., Hong Kong
\textsuperscript{4} Indiana University School of Public Health-Bloomington, USA

\textit{Abstract}

As the prevalence of type 2 diabetes mellitus and obesity increases worldwide, scientifically rigorous research is needed in this field to determine effective interventions for the prevention and treatment of these chronic diseases. In a recent study published in this journal, Zhou et al.\textsuperscript{1} conclude that metformin, a drug used for treatment of type 2 diabetes mellitus, can be used effectively for weight loss, and that this effect is even more pronounced in individuals who weigh more at baseline. Unfortunately, we believe these results to be due to the regression to the mean (RTM) phenomenon, which weakens the causal inference proposed in this study. The conclusions of Zhou et al. that metformin is an effective strategy for weight loss in individuals with type 2 diabetes mellitus are not substantiated due to the lack of a control group and failure to consider other factors that may have confounded these results.

\textbf{Key words:} statistics, metformin, body weight loss
Regression to the mean

Regression to the mean (RTM), though identified by Francis Galton over a century ago and well known to statisticians, is a statistical phenomenon that is commonly ignored or misunderstood and which leads to inappropriate causal conclusions in general, in biomedical research overall and in obesity research in particular. The phenomenon of RTM, if not addressed, can bias causal inference, and will occur in statistical analyses when examining 2 variables that are not perfectly correlated at a coefficient of 1.0. In the case of Zhou et al., these variables are BMI or body weight at 2 time-points during the study.

Zhou et al. focus the results and subsequent conclusions of this paper on the 2.4-kilogram weight loss observed over a period of 16 weeks. This level of weight change may not be due to the use of metformin alone. Without the use of a control group to compare changes in mean body weight over the course of the study, the authors cannot estimate the effect of metformin on weight loss separately from RTM or other factors which may lead to weight loss.

The second finding in this paper from Zhou et al. was the observation that individuals with higher body weight at baseline had lost more weight due to metformin by the end of the 16-week period. Even ignoring the tendency to increase type 1 errors – which the stepwise regression method the authors used demonstrates – this result is most likely due to RTM. By selecting a subgroup to analyze based on baseline measurements, there will be greater RTM than by observing the sample as a whole. If this subgroup of individuals has a higher baseline value for a given variable – body weight in this example – then it is expected that their second observation will also be high, but not as high as the first one, as they will have regressed towards the mean. The conclusions made by the authors state: “[w]e found that after metformin treatment, the proportion of obese patients decreased (p = 0.005).” The proportion of overweight patients also decreased, whereas the portion of normal weight patients increased, although not significantly is influenced by RTM. Furthermore, because weight change in this case is computed as the difference between weight at week 16 and baseline weight, it will be functionally associated with body mass index (BMI), which was used in the stepwise regression model. Because of both RTM and the formula of weight change itself, higher baseline weight will almost always be associated with greater weight loss.

This examination of participants at the extreme ends of the distribution will lead to greater RTM than if the sample had been examined in its entirety. This finding, which was the second major conclusion of the paper, is a result of RTM and therefore should not be considered evidence of any greater efficacy of metformin on weight loss in patients with obesity.

Conclusions

Without a control group, the causal statements and conclusions offered by Zhou et al. are not justified. Zhou et al. mention the lack of a control group as a limitation in the discussion, yet the conclusions are nevertheless causal and fail to address other potential confounding factors for the results observed in the study. The central finding that metformin treatment is associated with greater weight loss among those who were heaviest at baseline is predictable in view of RTM. Again, without a control group, this result cannot be taken as an indicator of differential efficacy as a function of baseline BMI. The article underscores the ongoing lack of understanding of RTM and its capacity to contribute to misleading conclusions in obesity research.

References