



HHS Public Access

Author manuscript

Pediatr Obes. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Pediatr Obes. 2016 December ; 11(6): e16–e17. doi:10.1111/ijpo.12030.

Inconsistencies and Inaccuracies in Reporting on Choice of Endpoints and of Statistical Results in RCT of Maternal Diet

Dwight W. Lewis Jr.⁽¹⁾, David A. Fields⁽²⁾, and David B. Allison⁽³⁾

⁽¹⁾Institute for Rural Health Research, College of Community Health Sciences, Tuscaloosa, AL, USA

⁽²⁾Department of Pediatrics, University of Oklahoma Health Sciences, Norman, OK, USA

⁽³⁾Office of Energetics and Nutrition Obesity Research Center, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA

Keywords

Trial Registry; Multiple Testing; Reproducibility

Dear Editor

We read Donnelly and colleagues' study (1) with interest, as neonatal body composition is a timely topic. Yet, we would like to highlight some concerns and probable errors, which we believe will allow readers to better interpret the report. These include: (i) inconsistency in primary endpoints between the study's publicly available trial registration and the published paper; (ii) apparent errors in the reporting of key findings; and (iii) neither adjusting for nor recognizing the plausibility that the one statistically significant finding among many significance tests was a type-I error.

Inconsistency between trial registry and manuscript

We would first like to highlight inconsistency in what is being claimed as endpoints in the trial's public registration (<http://www.isrctn.com/ISRCTN54392969>) and the published paper. The paper's methods section states "Primary outcome was birthweight and secondary outcome was gestational weight gain and glucose intolerance." Curiously, that is not reflected in the paper's title or abstract. More troublingly, the publicly available protocol for the study of interest does not list glucose intolerance as a primary nor as a secondary outcome. Moreover, the registered protocol does not mention many of the listed anthropometric measures as outcomes of interest. Inconsistencies such as these have been suggested to be problematic for medical research (2, 3). Perhaps this study is one of post-hoc analysis among a subgroup of participants from a randomized controlled trial (RCT). There is nothing wrong with post-hoc analyses, but they should be clearly described as such.

Corresponding author and Contact to request reprints: David B. Allison, PhD, University of Alabama at Birmingham, Office of Energetics, RPHB 140J, 1720 2nd Avenue South, Birmingham, AL 35294-0022, Phone: (205) 975-9169; Fax: (205) 975-7536; dallison@uab.edu.

Calculation and typographical errors

There are several inconsistencies and probable inaccuracies related to published findings in Donnelly and colleagues' manuscript. After careful observation of Table 3, we calculated that the summation of mean skin-folds for the intervention groups as 28.48 mm, while the control group's was 28.74 mm. This is different from what Donnelly and colleagues have calculated (intervention = 22.8 mm vs. control = 24.4 mm). There is also a small (2-tenths) difference between the control group's mean thigh circumference as reported in the abstract (16.6 cm) and Table 2 (16.4 cm), suggesting that one is a typographical error. Probably the most striking calculated error in the manuscript is related to the key finding of thigh circumference. The authors report that the between-group difference among neonates' thigh circumference is associated with a p-value of 0.04, but using the authors' data reported in their table, we obtain a p-value of 0.0116. If we use the data from their abstract, we obtain a p-value of 0.0004. That said, if we take the maximal possible rounding error for each group's mean and standard deviation, we would then calculate a p-value of 0.0488. Given these results it is difficult to discern what are typographical, calculation, or rounding errors, and it may be useful for Donnelly and colleagues to provide clarification.

Multiple testing issues

Uncontrolled and unaddressed multiple testing increases the odds of false positive findings (4). Based on Tables 2 and 3 alone, it appears that 14 significance tests were conducted. This does not take into account any subgroup testing or other testing that may have been conducted. Again, while there is nothing wrong with testing multiple hypotheses in post-hoc analyses, this should at minimum be mentioned in the text to alert readers to the issue. While there is no universal agreement on when a multiple testing 'correction' should be used, as an explanation of one limitations section checklist item of the 2010 CONSORT guidelines state "Authors should exercise special care when evaluating the results of trials with multiple comparisons. Such multiplicity arises from several interventions, outcome measures, time points, subgroup analyses, and other factors. In such circumstances, some statistically significant findings are likely to result from chance alone" (5). We believe the readers of the Donnelly et al article would have obtained a different impression had the authors made a clear statement about the matter such as:

"In post-hoc analyses involving at least 14 significance tests, we obtained one p-value which was significant at the nominal 0.05 alpha level. Although we will interpret it as a possible finding, readers may wish to note that had a Bonferroni correction for multiple testing been used, only p-values below $0.05/14 \approx 0.0036$ would have been significant. Thus, our one significant result is very plausibly a type 1 error."

Conclusion

Concerns about reproducibility in science have recently been elevated and the Committee on Publication Ethics suggests that inconsistencies and inaccuracies such as those identified here warrant correction (6). We believe that careful attention to such matters can help buttress the reliability of the scientific record.

Acknowledgments

Each author drafted some component of this letter to the editor, and all authors edited and approved all components of this letter.

Supported in part by NIH grants R25HL124208, R25DK099080, and P30DK056336. The opinions expressed are those of the authors and do not necessarily represent those of the NIH or any other organization.

Though not directly related to this project, Drs. Allison, Fields, and/or their universities have received funds from multiple food companies and pharmaceutical companies. Dr. Lewis reports no financial interests associated with the contents of this letter.

References

1. Donnelly JM, Walsh JM, Byrne J, Molloy EJ, McAuliffe FM. Impact of maternal diet on neonatal anthropometry: a randomized controlled trial. *Pediatr Obes*. 2015; 10:52–56. [PubMed: 24443392]
2. Altman DG, Moher D. Declaration of transparency for each research article. *BMJ*. 2013; 347:f4796. [PubMed: 23924655]
3. Ross, J. [accessed 29 January 2015] Trial registries, journals, and selective endpoint reporting: Still biased after all these years. *JAMA Internal Medicine Blog*. 2014. [WWW document]<http://internalmedicineblog.jamainternalmed.com/2014/04/21/trial-registries-journals-and-selective-endpoint-reporting-still-biased-after-all-these-years/>
4. Young SS, Bang H, Oktay K. Cereal-induced gender selection? Most likely a multiple testing false positive. *Proc Biol Sci*. 2009; 276:1211–1212. [PubMed: 19141426]
5. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010; 340:c869. [PubMed: 20332511]
6. Committee on Publication Ethics. [accessed 29 January 2015] Cope Code of Conduct for Journal Editors. 2011. [WWW document]http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf