



Published in final edited form as:

*Eur J Nutr.* 2020 September ; 59(6): 2813–2814. doi:10.1007/s00394-020-02287-0.

## **Within-group comparisons led to unsubstantiated conclusions in “Low-phytate wholegrain bread instead of high-phytate wholegrain bread in a total diet context did not improve iron status of healthy Swedish females: a 12-week, randomized, parallel-design intervention Study”**

**Colby J. Vorland<sup>1</sup>, Luis M. Mestre<sup>2</sup>, Sachintha S. Mendis<sup>3</sup>, Andrew W. Brown<sup>1</sup>**

<sup>1</sup>Department of Applied Health Science, Indiana University School of Public Health-Bloomington, Bloomington, Indiana, USA

<sup>2</sup>Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, Indiana, USA

<sup>3</sup>Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, Colorado, USA

---

Hoppe et al. recently published a study examining the effect of high and low phytate intake on iron status in a free-living population over 12 weeks (1), an important question in light of existing controlled feeding studies that were much more acute. However, we have concerns about the statistical analyses and interpretation that warrant a re-evaluation of the conclusions.

One of the purposes of a control group in parallel-group randomized controlled trials, such as that of Hoppe et al. (1), is to account for changes over time independent of an intervention to isolate the intervention effect. An error the authors repeatedly make is employing within-group instead of between-group comparisons to conclude differences between groups (the differences in nominal significance (DINS) error (2)). The DINS error has been shown in simulations to drastically inflate type 1 error rates (3, 4), and therefore will produce misleading conclusions. For example, in the abstract, they note “[i]n the high-phytate bread group (n = 31) there was no change in any of the iron status biomarkers...”, but “[i]n the low-phytate bread group (n = 24) there were significant decreases in both ferritin ... and total body iron”. They then conclude that “consumption of low-phytate wholegrain bread for 12 weeks resulted in a reduction of markers of iron status”. However, the appropriate between-group comparisons are reported in Table 1, and both ferritin and total body iron are reportedly not statistically significantly different between groups. Other examples of within-group comparisons are reported throughout the manuscript, including in tables and all figures. Instead, conclusions in the abstract, results, and discussion of this trial should focus on between-group differences between high- and low-phytate on each outcome, instead of interpreting each group separately.

We also note the reporting and interpretation of statistical significance have other limitations. All p-values should be fully reported per statistical reporting guidelines (5),

instead of p-values > 0.05 being reported only as “NS.” Appropriate statistical reporting permits back-calculation of between-group comparisons that are important for meta-analysis, appropriate evaluation of results (e.g., to be able to tell whether a NS p-value was actually 0.051 or 0.99), and to at least informally estimate family-wise error in the context of multiple comparisons (e.g., to estimate a Bonferroni correction when many p-values are reported). Further, conclusions should be stated so as not to accept the null. Currently, one of the concluding statements in the abstract states that “12 weeks of high-phytate wholegrain bread consumption had **no effect** on iron status” (1) [**emphasis added**]. However, it should be clarified that the authors failed to detect a statistically significant difference, rather than implying they are identical. This nuance reflects that the statistical tests used (null hypothesis significance testing) do not determine whether the null hypothesis is true, but rather whether the null hypothesis is an unlikely explanation for differences in the sample distributions under the assumptions of the test (6). Indeed, after dropouts and exclusions, the low-phytate group in Hoppe et al. (1) was 20% short of the estimated participants needed from their power calculation to reach 85% power. Underpowered tests are one possible explanation for failing to detect true differences at  $p < 0.05$ .

Finally, the authors reported randomizing participants into the high-phytate bread and dephytinized bread groups, where loss of participants “due to bread related reasons” was greater in the latter (16 vs 6) (Fisher’s exact test:  $p = 0.03$ ) (1). This differential loss between groups could potentially bias results, which the authors acknowledge. However, this is especially true if attributes of the treatment itself are the cause of dropout. Given the randomized design, we have just as much confidence in the treatment being the cause of dropout as we do in the treatment being the cause of any outcome. The authors report intention-to-treat (ITT) analyses for ferritin and total body iron, but only for within-group comparisons. They also report applying the “last observation carried forward” approach to handle missing data for these two ITT analyses, which has been extensively critiqued in statistical literature, e.g. (7). We suggest that between-group ITT comparisons also be fully reported. The authors may also be interested in sensitivity analyses when facing challenging missing data problems using different approaches to explore the robustness of their results.

We again commend Hoppe et al. for asking an important question regarding phytate and iron status in a randomized design. Our concerns expressed herein are mostly addressable by a reanalysis or reframing of their existing results, but at present the results are incomplete and the conclusions misleading in the context of a between-group randomized design. We request that the authors reanalyze their data to account for these statistical concerns and fully and transparently report their results. We offer our assistance in the reanalysis if requested.

## Acknowledgments

### Funding

Supported in part by the Gordon and Betty Moore Foundation and NIH grant R25HL124208. The opinions expressed are those of the authors and do not necessarily represent those of the NIH or any other organization.

### Disclosures

In the last 12 months, Ms. Mendis has been involved in research through Colorado State University for a project funded by NIFA-AFRI, USA. She also received travel grants and registration fee waivers from Indiana University

School of Public Health and Food Systems-Colorado State University Extension. Dr. Brown has received travel expenses from University of Louisville; speaking fees from Purdue University; and grants through his institution from Dairy Management, Inc., National Cattlemen's Beef Association, and NIH/NHLBI. He has been involved in research for which his institution or colleagues have received grants from Gordon and Betty Moore Foundation, NIH/NHLBI, NIH/NIA, NIH/NIDDK, and Sloan Foundation. Other authors report no disclosures.

## References

1. Hoppe M, Ross AB, Svelander C, Sandberg A-S, Hulthén L. Low-phytate wholegrain bread instead of high-phytate wholegrain bread in a total diet context did not improve iron status of healthy Swedish females: a 12-week, randomized, parallel-design intervention study. *European journal of nutrition*. 2019;58(2):853–64. [PubMed: 29796932]
2. Allison DB, Brown AW, George BJ, Kaiser KA. Reproducibility: A tragedy of errors. *Nature*. 2016;530(7588):27–9. [PubMed: 26842041]
3. Bland JM, Altman DG. Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials*. 2011;12(1):264. [PubMed: 22192231]
4. 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. [PubMed: 26354536]
5. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the “Statistical Analyses and Methods in the Published Literature” or the SAMPL Guidelines”. *Int J Nurs Stud*. 2015;52(1):5–9. [PubMed: 25441757]
6. Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016;70(2):129–33.
7. Lachin JM. Fallacies of last observation carried forward analyses. *Clinical trials*. 2016;13(2):161–8. [PubMed: 26400875]