Statistical thinking in simulation design: a continuing conversation on the balancing intercept problem

Boyi Guo, MS, PhD1; Linzi Li, MPH, MSPH2; Jacqueline Rudolph, PhD3

1. Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD
2. Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA
3. Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

**Corresponding Author**:  
Boyi Guo, PhD, MS

Department of Biostatistics

Bloomberg School of Public Health

Johns Hopkins University, Baltimore, MD

bguo6@jhu.edu

Abstract word count: 121

Non-abstract word count: 1848

1 Figure

# Abstract

Epidemiologists have a growing interest in employing computational approaches to solve analytic problems, with simulation being arguably the most accessible among all approaches. While previous literature discussed the utility of simulation and demonstrated how to carry out them, few have focused on connecting underlying statistical concepts to these simulation approaches, creating gaps between theory and application. Based on the recent series of discussions on the *balancing intercept*, we explain the growing complexity when generalizing the balancing intercept to a wider class of simulations and revise the closed-form equation for the balancing intercept under assumptions. The discussion can broadly inform the future design of more complex simulation and emphasize the importance of applying statistical thinking in the new era of computational science.

`

# Reference

1. Rudolph, J.E., M.P. Fox, and A.I. Naimi, *Simulation as a Tool for Teaching and Learning Epidemiologic Methods.* Am J Epidemiol, 2021. **190**(5): p. 900-907.

2. Fox, M.P., et al., *Illustrating How to Simulate Data From Directed Acyclic Graphs to Understand Epidemiologic Concepts.* Am J Epidemiol, 2022. **191**(7): p. 1300-1306.

3. Rudolph, J.E., et al., *SIMULATION IN PRACTICE: THE BALANCING INTERCEPT.* American Journal of Epidemiology, 2021. **190**(8): p. 1696-1698.

4. Robertson, S.E., J.A. Steingrimsson, and I.J. Dahabreh, *Using Numerical Methods to Design Simulations: Revisiting the Balancing Intercept.* American Journal of Epidemiology, 2022. **191**(7): p. 1283-1289.

5. Durrett, R., *Probability: theory and examples*. Vol. 49. 2019: Cambridge university press.

6. Zivich, P.N. and R.K. Ross, *RE: “USING NUMERICAL METHODS TO DESIGN SIMULATIONS: REVISITING THE BALANCING INTERCEPT”.* American Journal of Epidemiology, 2022.

7. Casella, G. and R.L. Berger, *Statistical inference*. 2021: Cengage Learning.

8. Te Grotenhuis, M., et al., *When size matters: advantages of weighted effect coding in observational studies.* International Journal of Public Health, 2017. **62**(1): p. 163-167.

A picture containing text, number, font, screenshot

Description automatically generated  
**Figure 1**: **The closed-form estimation of the balancing intercept controls the marginal mean at the target level for log-normal data generating models with different covariates settings.** The bias, defined as the empirical mean of the simulated outcome minus the targeted marginal mean of the outcome, holds at 0 for log normal data generating models of four different risk ratio magnitude of the covaraites and four different distribution for the covariates, including (**A**) a Bernoulli distribution with probability 0.8, (**B**) a gamma distribution with shape 1 and rate 1.5, (**C**) a standard normal distribution, and (**D**) a continuous uniform distribution bounded between -1 and 3.

Ethics approval and consent to participate  
Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Conflict of Interest

None declared.