Linking missing data to study outcomes using multiple imputations

Dear Editor:

Immigrants arriving in Canada are, on average, healthier than their non-immigrant counterparts, but are at an increased risk for developing poor health outcomes over time compared to non-immigrants; a phenomenon known as the 'healthy immigrant effect'.¹ However, little is known about the health outcomes, such as obesity in children of immigrants. A recent article analyzed data from the Canadian Community Health Survey to examine the differences in body mass index (BMI) and prevalence of being overweight/obese among immigrant and non-immigrant youth.² According to the results, immigrant youth had a lower zBMI and a lower prevalence of being overweight/obese relative to non-immigrant youth, but length of time in Canada was associated with higher zBMI scores and increased odds of being overweight/obese. The authors also found a positive association with energy expenditure and zBMI, which they acknowledged is contrary to previously published literature. The authors attribute this to a lack of robustness in the measure, but it could also be due to missing data.

The authors stated that one of the limitations of the study was a large percentage of missing data. In fact, over a third (approximately 33.6%) of the samples had one or more missing responses on study variables. Excluding missing data presents several problems, such as reduced power as well as threatened validity of statistical inference.³ To mitigate this limitation, the authors used multiple imputation (MI), which refers to the practice of 'filling in' missing data with plausible data by using an algorithm on SPSS that is based on linear regression.⁴ However, there was no indication of the mechanism responsible for the missing observational points nor a description of the study variables that were missing.

There are three mechanisms through which missing data can arise: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Multiple imputation

is acceptable depending on the category of missing data.⁴ However, it is not possible to distinguish between MAR and MNAR using observed data.³ In these instances, bias can occur when data are MNAR and this can only be addressed by sensitivity analysis which will examine the effect of different assumptions about the missing data mechanism.³

It may be worthwhile for the authors to consider whether the missing data were differential or non-differential between the immigrant and non-immigrant populations, as this can have an impact on the conclusions drawn from the data. If equal proportions of missing data were reported for both groups, then the underlying assumptions for multiple imputations are more likely to be valid. However, if data are differentially missing, such as an increased proportion of non-response among immigrants, then the assumptions of multiple imputation may be invalid (i.e., data are not missing at random).³ It is important to account for systematic differences between missing values and observed values between the two groups. If multiple imputation is used, it may provide misleading results, which may be what led to the paradoxical conclusions in the study in question.

Khadija Ibrahim, MSc, MPH

Division of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, E-mail: khadija.ibrahims@gmail.com

doi: 10.17269/CJPH.106.4914

REFERENCES

1. McDonald JT, Kennedy S. Insights into the 'healthy immigrant effect': Health status and health service use of immigrants to Canada. *Soc Sci Med* 2004;59(8):1613–27. doi: 10.1016/j.socscimed.2004.02.004.
2. Wahi G, Boyle MH, Morrison KM, Georgiades K. Body mass index among immigrant and non-immigrant youth: Evidence from the Canadian Community Health Survey. *Can J Public Health* 2013;105(4):e239–e244.
3. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ* 2009;338:b2393.
4. Royston P. Multiple imputation of missing values. *Stata J* 2004;4:227–41.