## Authors' reply

Singapore Med J 2019; 60(2): 104-105 https://doi.org/10.11622/smedj.2019019

Dear Sir,

We thank the writer for sharing his comments[1] in response to our article.[2] The writer suggested that the odds ratios (ORs) predicting diabetes mellitus for some ethnic groups were inflated because of "*separation in logistic regression*",[3] which occurs when one or more of a model's covariates, in this case ethnicity, perfectly predicts some binary outcome. Numerically, there are two types of separation: with complete separation, the outcome of each subject in the data set can be perfectly predicted, while with quasi-complete separation, this is possible for only a subset of the subjects. The simplest example is a 2 × 2 table of Y and X with an empty cell. In the case of a binary covariate X, complete separation corresponds to the case in which only the two opposing diagonal cells of the table contain data; in such circumstances, Y can be perfectly predicted by X for all the observations in the data.[4] Based on these examples, we did not find the writer's reason to be applicable to our data, since a simple cross-tabulation between ethnicity and diabetes mellitus did not reveal any empty cells.

Secondly, possible collinearity between obesity and ethnicity was suggested. Possible collinearity problems between predictors can usually be determined by obtaining the variance inflation factor (VIF).[5] We found no significant multicollinearity effect between obesity and ethnic group (VIF values ranging from 1.00 to 1.01). The proportion of obesity in our sample was not significantly different among the three ethnic groups (Chinese 42%, Malay 45%, Indian 33%; $\chi^2 = 0.2425$, $p = 0.886$).

Thirdly, we would like to thank the writer for pointing out that waist-to-height ratio, waist circumference and waist-hip ratio are better predictors compared to body mass index for cardiovascular risk mortality. We did not include these predictors in the current study but will keep this in mind for future studies.

Lastly, the writer suggested that using 'atypical low' as the reference category may produce significant results. However, we feel that 'atypical low' (n = 9) should not be used as the reference category due its small sample size. When choosing a reference group, sample size in each category is an important factor to consider. The smallest group should be avoided as "*standard errors of coefficients for other categories will be inflated due to the small sample size in the reference group*".[6] 'Typical high' (n = 30) was chosen to be the reference group due to clinical importance, as long-stay patients are usually on a high dosage of antipsychotic medication and were thus the group of interest in this study. Even though the 'atypical and typical (combined) high' group (n = 38) had the highest sample size, it was not chosen, as it is a mixed category rather than a well-defined group. A guideline for choosing the reference group[7] recommended selecting a clearly defined group. Nevertheless, as suggested by the author, we reanalysed the data using 'atypical low' as a reference category. The results remained the same, with no significant association between medication type and diabetes mellitus. We found that the results for 'typical high' (OR 1.41, p = 0.77), 'typical low' (OR 1.75, p = 0.66) and 'atypical and typical (combined) high' (OR 0.81, p = 0.87) were not significant when compared to those for 'atypical low'.

Yours sincerely,

Saleha Shafie[1], Siau Pheng Lee[1], Samantha Bee Cheng Ong[2], Peizhi Wang[1], Esmond Seow[1], Hui Lin Ong[1], Siow Ann Chong[1], Mythily Subramaniam[1]

[1]Research Division, [2]Nursing Administration, Institute of Mental Health, Singapore. saleha_shafie@imh.com.sg

## References

1. Afifi M. Comment on: Prevalence and correlates of diabetes mellitus and dyslipidaemia in a long-stay inpatient schizophrenia population in Singapore. Singapore Med J 2019; 60:104.
2. Shafie S, Lee SP, Ong SBC, et al. Prevalence and correlates of diabetes mellitus and dyslipidaemia in a long-stay inpatient schizophrenia population in Singapore. Singapore Med J 2018; 59:465-71.
3. Mansournia MA, Geroldinger A, Greenland S, Heinze G. Separation in logistic regression: causes, consequences, and control. Am J Epidemiol 2018; 187:864-70.
4. Zorn C. A solution to separation in binary response models. Political Anal 2005; 13:157-70.
5. Midi H, Sarkar SK, Rana S. Collinearity diagnostic of binary logistic regression model. J Interdiscip Math 2010; 13:253-67.
6. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. New York, NY: Springer, 2005.
7. Hardy MA. Regression with Dummy Variables. Newbury Park, CA: Sage, 1993.