The association between the neighbourhood characteristics and depression: was the regression model satisfactory?

The study by Generaal and colleagues is noteworthy because of its large sample size and its objective and intelligent measurements of the myriad neighbourhood characteristics. However, we would like to draw the readers’ attention to the regression models. The Supplementary Table 1, with the bivariate correlations among the independent variables, showed modest to strong correlations between large numbers of variables. Therefore, multicollinearity was present. Under such circumstances, it is advisable to do variance inflation factor (VIF) estimation. VIF of more than ten suggests multicollinearity is a significant problem. Independent variables with VIF more than ten should have been removed from the model. The other option is to carry out principal component analysis of highly correlated independent variables. As the authors have not undertaken either of these corrections, significant multicollinearity might have affected the magnitude of the standardised regression coefficients, their standard errors and the P-values. These could potentially result in unreliable interpretations. The authors could have added the proportion of variance (R²) in the dependent variable (depression prevalence/severity) explained by the independent variables (neighbourhood characteristics) because R² is not affected by multicollinearity. Additionally, R² would have given an idea about the goodness of fit of the regression models.

The severity of depression (i.e. the dependent variable) had skewed distributions in five out of the seven cohorts. We agree, with a large sample size linear regression analysis could be done, even with a non-parametric dependent variable. However, the ordinary least square estimations should have been carried out to demonstrate statistical robustness of the regression analysis. In the case of non-normality of the ordinary least square, bootstrapping is an alternative.

Because of these limitations, we would be cautious while interpreting the results of the regression analysis undertaken to examine the association between the neighbourhood characteristics and severity of depression.