Correspondence

Edited by Kiriakos Xenitidis and Colin Campbell

Contents

- The association between the neighbourhood characteristics and depression: was the regression model satisfactory?
- Authors' reply

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.[1] 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.[2] 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.[3] The authors could have added the proportion of variance (R^2) in the dependent variable (depression prevalence/severity) explained by the independent variables (neighbourhood characteristics) because R^2 is not affected by multicollinearity. Additionally, R^2 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.[4]

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.

1 Generaal E, Hoogendijk EO, Stam M, Henke CE, Rutters F, Oosterman M, et al. Neighbourhood characteristics and prevalence and severity of depression: pooled analysis of eight Dutch cohort studies. *Br J Psychiatry* 2019; **215**: 468–75.

2 Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiol* 2016; **6**: 227–46.

3 Kim JH. Multicollinearity and misleading statistical results. *Korean J Anesthesiol* 2019; **72**: 558–69.

4 Hubert M, Rousseeuw PJ, Aelst SV. Inconsistency of resampling algorithms for high-breakdown regression estimators and a new algorithm. *J Amer Stat Assoc* 2002; **97**: 151–3.

Abhishek Ghosh, Assistant Professor, Department of Psychiatry, Post Graduate Institute of Medicne Education and Research, India; **Natarajan Varadharajan**, Senior Resident, Department of Psychiatry, Post Graduate Institute of Medicine Education and Research, India. Email: ghoshabhishek12@gmail.com

Authors' reply

We thank Ghosh and Varadharajan for the attention they have given to our paper 'Neighbourhood characteristics and prevalence and severity of depression: pooled analysis of eight Dutch cohort studies'.[1] They raised some concerns about the multilevel regression analyses conducted for each of the eight contributing studies that were pooled in a meta-analysis.[2]

First, Ghosh and Varadharajan point out the potential issue of multicollinearity when highly correlated variables are entered within one regression model. However, this appears to be based on a misunderstanding. As indicated in the Method, all models run in our paper are univariable analyses performed for each neighbourhood variable in separate regression models. It is indeed true that different neighbourhood characteristics are moderately to strongly correlated. That is exactly the reason why we decided not to run multivariate analyses in which all environmental characteristics are entered within one model. By analysing them separately, we prevented the risk of multicollinearity and we provided better insights into which environmental characteristics are and are not associated with depression. We believe epidemiological studies that consider and compare multiple environmental characteristics, instead of focusing on only one characteristic, are very much needed as these give us a fuller understanding of the exposome relevance for mental health.

Second, Ghosh and Varadharajan indicated that non-normal distributions of depressive symptoms are not ideal for regression analyses. Indeed, in some of our cohorts the depressive symptom score was a bit skewed. However, we do not believe this has had impact on our overall results and conclusion. It is important to point out that the continuous depressive symptom score was only an outcome used in secondary sensitivity analyses. Our primary outcome measure was a dichotomous indicator of yes/no reporting significant depressive symptoms. Findings of secondary sensitivity analyses with a continuous outcome were very similar to that of primary analyses with a dichotomous outcome. In addition, as the skewness of the depressive symptom score was an issue in some but not in other cohorts, if this would have an impact, one would expect to see differences in associations across studies. However, our heterogeneity analyses showed in fact rather low heterogeneity in most results across the eight cohorts. So, we feel that also this issue did not impact on our results, which indicate – in a large-scale pooled analysis – that urbanisation and various socioeconomic, physical and social neighbourhood characteristics are associated with depression.

1 Generaal E, Hoogendijk EO, Stam M, Henke CE, Rutters F, Oosterman M, et al. Neighbourhood characteristics and prevalence and severity of depression: pooled analysis of eight Dutch cohort studies. *Br J Psychiatry* 2019; **215**: 468–75.

2 Ghosh A, Varadharajan N. The association between the neighbourhood characteristics and depression: was the regression model satisfactory? *Br J Psychiatry* 2020; **216**:235.

Ellen Generaal, PhD, Postdoc researcher, Department of Psychiatry, Amsterdam UMC, the Netherlands; **Brenda W.J.H. Penninx**, Professor, Department of Psychiatry, Amsterdam UMC, the Netherlands. Email: egeneraal@ggd.amsterdam.nl