Appendix 3 (as supplied by the authors): Detailed Statistical Analysis

Regression Analysis

We first performed regression analyses to examine the determinants of the three main mental health outcomes, including psychological distress scale, lifetime suicidal ideation and lifetime suicide plan. We assessed the differences in psychological distress and suicidal behaviour between non-Indigenous and Indigenous populations living off-reserve in Canada using the following equation:

$$Y_i = \alpha_0 + \alpha_1 Indigenous_i + \beta X_i + \mu_i. \quad (1)$$

In Equation 1, Y_i denotes the outcome variable for person i, α_0 is the intercept, α_1 is the coefficient on the dummy variable for Indigenous population, X_i is a vector of other control variables (demographic, socioeconomic and geographic factors) and μ_i is the error term. We performed the ordinary least squares and logistic models to investigate the effect of determinants on the continuous psychological distress, and two binary suicidal behaviours, respectively.

The Blinder–Oaxaca Decomposition

We employed the Blinder-Oaxaca (BO) decomposition method[1,2] to understand the contribution of each factor to the overall differences in the three mental health outcomes between non-Indigenous and Indigenous populations. The BO technique enabled us to decompose the observed gaps between non-Indigenous and Indigenous populations into explained and unexplained components. The explained (endowment) component captures the part of the difference in a given outcome explained by differences between groups in the level of observed characteristics (determinants) that were assessed (i.e., demographic, socioeconomic and geographic variables). The unexplained (response or return) component captures the portion attributable to differences in the effects of these characteristics and unobserved determinants on the outcome of interest in non-Indigenous and Indigenous populations. The absolute value of the explained component for income factor, for example, can determine how much the gap in the mean of psychological distress levels between non-Indigenous and Indigenous populations would reduce if the income of Indigenous population is set to be at the similar level of non-Indigenous Canadian, *ceteris paribus.* [1–3]

Appendix to: Hajizadeh M, Hu M, Asada Y, et al. Explaining the gaps in psychological distress and suicidal behaviours between non-Indigenous and Indigenous adults living off-reserve in Canada: a cross-sectional study. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200177.

We applied the two-fold BO decomposition and assume Indigenous population have comparative disadvantages than non-Indigenous population in relation to the three mental health outcomes being assessed.[1] We considered that the psychological issues among Indigenous and non-Indigenous adults are related to vectors of demographic, socioeconomic and geographic factors for non-Indigenous and Indigenous populations, respectively, as follows:

$$Y_{Ni} = \alpha_N + X_{Ni}\beta_N + \mu_{Ni} \qquad \text{and} \quad Y_{Ii} = \alpha_I + X_{Ii}\beta_I + \mu_{Ii}. \qquad (2)$$

Where α_I and α_N are intercepts. β_I and β_N are the coefficients indicating the effect of the observable characteristics on mental health outcomes, and μ_{Ii} and μ_{Ni} are error terms with zero expected value. If we take the expectations of non-Indigenous and Indigenous populations' mental health outcomes and subtract the expected values between the two groups, we can get Equation 3.[2]

$$Diff = \bar{Y}_N - \bar{Y}_I = \alpha_N + \bar{X}_N\beta_N - \alpha_I - \bar{X}_I\beta_I. \qquad (3)$$

Adding and subtracting the term $\bar{X}_I\beta_N$ in Equation 3 leads to:

$$Diff = (\bar{X}_N - \bar{X}_I)\beta_N + \bar{X}_I(\beta_N - \beta_I) + (\alpha_N - \alpha_I). \qquad (4)$$

The first component in Equation 4, $(\bar{X}_N - \bar{X}_I)\,\beta_N$, is the explained component and represents the difference in the given mental health outcome between two groups that is explained by variation in the mean level of observable characteristics. The second component, $\bar{X}_I(\beta_N - \beta_I) + (\alpha_N - \alpha_I)$, indicates the unexplained component and represents the difference in the outcome of interest that is due to differences in returns to observable and unobserved characteristics on the outcome of interest between two groups.

We used the BO decomposition for a linear model to assess the differences in the mean psychological distress. The extended version of the BO technique for a logistic model [3] was

[1] Two-fold OB decomposition assumes there is a potential disadvantage (negative discrimination) on Indigenous' psychological outcomes and there is no advantage (positive discrimination) for non-Indigenous population.

[2] We assume there is a comparative disadvantage against Indigenous population's psychological health, comparing to non-Indigenous population. Since the average levels of mental health outcome variables for Indigenous population are higher than their non-Indigenous counterparts, the difference in Equation 3 is negative.

Appendix to: Hajizadeh M, Hu M, Asada Y, et al. Explaining the gaps in psychological distress and suicidal behaviours between non-Indigenous and Indigenous adults living off-reserve in Canada: a cross-sectional study. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200177.

used to examine the difference in the rates of suicidal ideation and suicide plans. Past studies have highlighted certain issues related to the choice of the reference group when dummy or categorical variables used in the decomposition analysis.[4–6] Specifically, the estimated coefficients are sensitive to the choice of the base category when using a categorical variable in the model; thus, the decomposition results vary by altering the base group. To overcome this problem, as suggested by Gardeazabal and Ugidos[7] and Yun[5], we applied the deviation contrast transformation to each set of categorical variables and restricted the coefficients for the same set of categorical variables sum up to zero (normalize). The results of this approach are equal to performing a series of decompositions by changing the reference category one by one and calculating the averages of the results.[5]

References

1. Oaxaca R. Male-Female Wage Differentials in Urban Labor Markets. *Int Econ Rev (Philadelphia)*. Blackwell Publishing for the Economics Department of the University of Pennsylvania and Institute of Social and Economic Research -- Osaka University; 1973;**14**(3):693–709.
2. Blinder AS. Wage Discrimination: Reduced Form and Structural Estimates. *J Hum Resour*. University of Wisconsin Press; 1973;**8**(4):436–455.
3. Powers DA, Yoshioka H, Yun MS. Mvdcmp: Multivariate decomposition for nonlinear response models. *Stata J*. 2011;**11**(4):556–576.
4. Horrace WC, Oaxaca RL. Inter-Industry Wage Differentials and the Gender Wage Gap: An Identification Problem. *Ind Labor Relations Rev*. 2001 Apr;**54**(3):611.
5. Yun M-S. A Simple Solution to the Identification Problem in Detailed Wage Decompositions. *Econ Inq*. Wiley/Blackwell (10.1111); 2005 Oct;**43**(4):766–772.
6. Yun M-S. Hypothesis tests when decomposing differences in the first moment. *J Econ Soc Meas*. North-Holland; 2005 Jan;**30**(4):295–304.
7. Gardeazabal J, Ugidos A. More on identification in detailed wage decompositions. *Rev Econ Stat*. 2004;**86**:1034–1036.

Appendix to: Hajizadeh M, Hu M, Asada Y, et al. Explaining the gaps in psychological distress and suicidal behaviours between non-Indigenous and Indigenous adults living off-reserve in Canada: a cross-sectional study. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200177.