

Example dataset for disclosure risk assessment

- CASE ID: unique number for each case
- Province or Territory
- Strata: a geographic unit for sampling
- Mental Health: 1 = has had diagnosed mental health issues, 2 = has not
- Substance Abuse: 1 = has been treated for substance abuse issues, 2 = has not
- Family Structure: 1 = lives alone, 2 = one family, 3 = blended family, 4 = multigenerational family, 5 = other
- Education: number of years of education
- Income: monthly income in dollars
- Employment: 1 = employed full time, 2 = employed part time, 3 = employed on call, 4 = unemployed
- Ethnic Origin: 1 = First Nations, 2 = Métis, 3 = Inuit, 4 = South Asian, 5 = West European, 6 = Latin American, 7 = other
- Age: in years and months at time of survey (derived from survey date and birth date)
- Gender: 1 = female, 2 = male, 3 = neither, 4 = both, 5 = other

Classify each variable in the dataset as:

- a. re-identification
- b. sensitive
- c. no risk
- d. indeterminate

Are there combinations of variables that increase disclosure risk?