Mental health among the sugar cane industry farmers and non-farmers in Peru: a cross-sectional study on occupational health

**Supplementary 1:** Sample size and power analysis

**Sample size for the main study:**

On the main study, to determine the sample size required to detect differences in eGFR of -5 ml/min/1.73m² or higher (1), we used a significance level of 5% (α = 0.05) for a 2-tailed test with a statistical power of 80% (1-β = 0.80) and an estimated variance of 289 (2). With this information, assuming that the population is infinite, 81 people per group (exposed and unexposed) were obtained. Assuming a 20% loss to follow-up, a sample size of 97 was obtained at a ratio of 1:1 per group, 97 farmers to 97 non-farmers. However, to increase the power of this study, the inclusion ratio for this study was 2:1, for a total of 291 participants.

Also, random sampling stratified by age (18-30, 31-45, 46-60 years) and work activity was used in the main study. The list of workers was used as a sampling frame in our database. The farmers (exposed group) who met the study’s inclusion and exclusion criteria were chosen, and non-farmer workers with similar characteristics (age) to each farmer.

**Power analysis for the current manuscript:**

**Two-Sample T-Test Power Analysis: Numeric Results for Two-Sample T-Test**

Null Hypothesis: Mean1=Mean2. Alternative Hypothesis: Mean1≠Mean2

The standard deviations were assumed to be unknown and unequal.

<table>
<thead>
<tr>
<th>Allocation</th>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.00</td>
<td>175</td>
<td>106</td>
<td>0.606</td>
<td>0.05</td>
<td>0.00</td>
<td>10.7</td>
<td>9.5</td>
<td>0.3</td>
<td>0.2</td>
</tr>
</tbody>
</table>

**Report Definitions**

- Power is the probability of rejecting a false null hypothesis. Power should be close to one.
- N1 and N2 are the number of items sampled from each population. To conserve resources, they should be small.
- Alpha is the probability of rejecting a true null hypothesis. It should be small.
- Beta is the probability of accepting a false null hypothesis. It should be small.
- Mean1 is the mean of populations 1 and 2 under the null hypothesis of equality.
Mean2 is the mean of population 2 under the alternative hypothesis. The mean of population 1 is unchanged.

S1 and S2 are the population standard deviations. They represent the variability in the populations.

Summary Statements
Group sample sizes of 175 and 106 achieve 100% power to detect a difference of 1.2 between the null hypothesis that both group means are 10.7 and the alternative hypothesis that the mean of group 2 is 9.5 with estimated group standard deviations of 0.3 and 0.2 and with a significance level (alpha) of 0.05000 using a two-sided two-sample t-test.

References:
Supplementary 2: Flowchart based on the sample agreed upon

A total of 1000 workers participated in the study. We obtained a sample size of 291 workers from that total. Only 95.6% (281 workers) agreed to take part in the study.
Supplementary 3: Assumption’s evaluation

Examine equidispersion

(i) Poisson goodness-of-fit test

```
poisgof
```

```
Deviance goodness-of-fit = 538.2604
Prob > chi2 (279) = 0.0000
```

```
Pearson goodness-of-fit = 575.6407
Prob > chi2 (279) = 0.0000
```

Conclusion: The Poisson goodness-of-fit test results indicate (p-value 0.05) that the Poisson model is inappropriate. Similarly, when the deviance was divided by the number of observations, the value was > 1, indicating overdispersion. Both results show that Negative Binomial Regression should be used instead of Poisson Regression.

(ii) The alpha parameter for overdispersion

```
xi: nbreg ghq12_num cond_labor
```

```
Negative binomial regression
Dispersion = mean
Log likelihood = -511.89235
```

```
   | Coef.  | Std. Err. |    z |  P>|z|     |  [95% Conf. Interval]           |
-------------|---------|----------|------|--------|---------------------------------|
ghq12_num    | .8329231| .1196618 | 6.96 | 0.000  | .5953908 to .107456             |
            | .2905351| .0831008 | 3.50 | 0.000  | .1276606 to .4534097            |
            | .7751364| .2216944 | -3.48 | 0.000  | -1.209649 to -.3406233          |
            | .460641 | .1021215 | 4.50 | 0.000  | .2593018 to .662268             |
```

Conclusion: The overdispersion alpha parameter test results show that the alpha is significantly different from zero, reinforcing the position that the Poisson distribution is inappropriate.