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.

Allocation

Power	N1	N2	Ratio	Alpha	Beta	Mean1	Mean2	S1	S2
1.00	175	106	0.606	0.05	0.00	10.7	9.5	0.3	0.2

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.

29   − Mean2 is the mean of population 2 under the alternative hypothesis. The mean of
30   population 1 is unchanged.
31   − S1 and S2 are the population standard deviations. They represent the variability in the
32   populations.

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

38   **References:**

39   1. González-Quiroz, M., Camacho, A., Faber, D. *et al.* Rationale, description and baseline
40   findings of a community-based prospective cohort study of kidney function amongst the
41   young rural population of Northwest Nicaragua. *BMC Nephrol.* 2017;**18**,16.

42   2. Laws RL, Brooks DR, Amador JJ, Weiner DE, Kaufman JS, Ramírez-Rubio O, et al.
43   Biomarkers of Kidney Injury among Nicaraguan Sugarcane Workers. Am J Kidney Dis.
44   2016;67(2):209–17.

45   3. Hounsome B, Edwards RT, Hounsome N, Edwards-Jones G. Psychological morbidity of
46   farmers and non-farming population: Results from a uk survey. Community Ment Health
47   J. 2012;48(4):503–10.

48   **Supplementary 2:** Flowchart based on the sample agreed upon

49   A total of 1000 workers participated in the study. We obtained a sample size of 291 workers
50   from that total. Only 95.6% (281 workers) agreed to take part in the study.
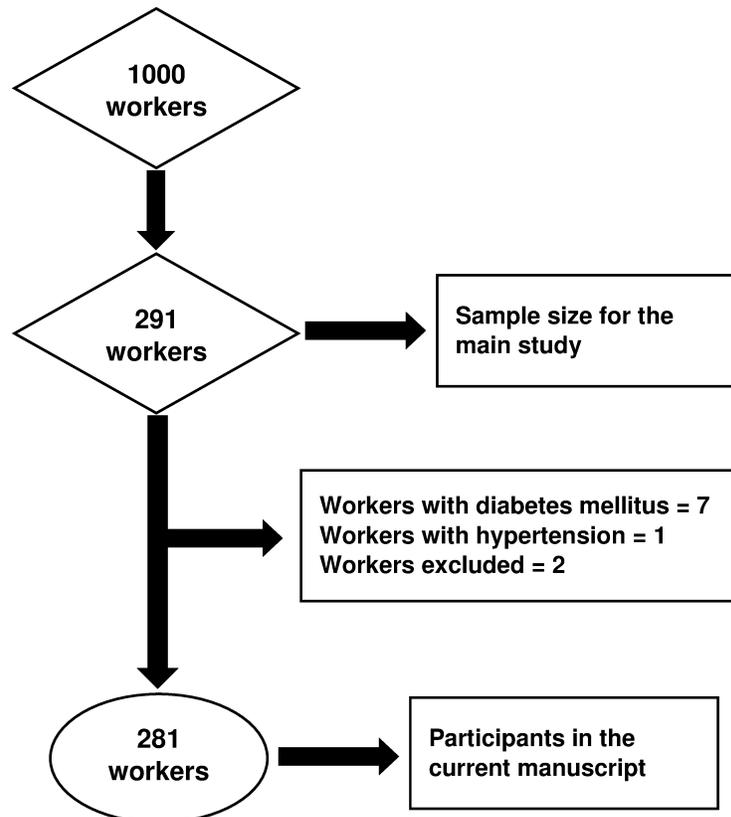
51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67   **Supplementary 3:** Assumption's evaluation

68   Examine equidispersion

69   (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
display 538.2604/279
1.9292487
```

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

79   (ii) The alpha parameter for overdispersion

```
xi:nbreg ghq12_num cond_labor

Negative binomial regression                    Number of obs   =        281
                                                LR chi2(1)      =      43.74
Dispersion     = mean                           Prob > chi2     =     0.0000
Log likelihood = -511.89239                     Pseudo R2       =     0.0410


------------------------------------------------------------------------------
   ghq12_num |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  cond_labor |   .8329231   .1196616     6.96   0.000     .5983908    1.067456
       _cons |   .2905351   .0831008     3.50   0.000     .1276606    .4534097
-------------+----------------------------------------------------------------
    /lnalpha |  -.7751364   .2216944                     -1.209649   -.3406233
-------------+----------------------------------------------------------------
       alpha |    .460641   .1021215                      .2983018    .7113268
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 48.86             Prob >= chibar2 = 0.000
```

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