Supplementary material. Statistical methods

To calculate the sample size, an expected seroprevalence of 30% (p = 0.30), (q = 1-0.30 = 0.70) was taken with a marginal sampling error of ± 2% (δ = 0.02) and a confidence level = 95 % (α = 0.05, Zα = 1.96). An additional 10% was included in the sample size considering potential losses. The sample size was divided for each city following a proportional allocation criteria [1], according to the National Health Workers Database [2].

The research team had access to the epidemiological reports, allowing the classification of the IPS according to the number of COVID-19 treated in each hospital or clinic. IPS who gathered the 80% of the cases in the city were invited to participate in the study. In case the IPS was not interested in participating, the next institution on the list was invited.

We asked to the Human Resources Office for a list of the health care workers for each service (A&E, hospitalization, surgery rooms, laboratory, radiology, general services, security, catering, cleaning, administrative, etc.) Using a random number table in Excel® we selected the potential participants for each service in the study [3]. Potential participants were invited to take part in the research. In case that HW were not able to participate, we follow to the next potential participant in the list.

In some cases, the number of volunteers in each institution exceed the minimum sample size. This doesn´t affect the power of the study since a greater sample increased the relative precision, passing from 7% to 5% [4,5].

Poisson analysis description

A robust *Poisson* regression was performed since the expected prevalence was greater than 20%. To avoid overestimating the effect, we used the prevalence ratio. In this sense, the *Poisson* regression avoid an overestimation bias [6]. In addition, the variance was robust, since it does not comply with the assumptions of normality of variables of other generalized linear models.

The used model was analyzed in R V and the syntax used was:

Script for Poisson model

1

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance
placed on this supplemental material which has been supplied by the author(s)

Occup Environ Med

```
library(readxl)
library(Epi)
library(foreign)

grafi      <-      read_excel("D:/R      STUDIO/INS/2020/RESULTADOS      GLOBAL      VS
ENCUESTAS/modelo.xls",
           sheet = "Hoja2", convert.factors=F)
glm.5 <- glm (RESULTADO ~ edadcod2 + Sexo +
        EstratoSocial +  TipodeSangreyRh +  NumeroTrabajos +
        ServiciosTrabaja+  Comorbilities+ fumar+  HorasTrabajadas,
       family = poisson(link = "log"), data = grafi).
summary(glm.5)

#### PR crude
PR_IC <- Confint (glm.5, level=0.95, type="LR", exponentiate=TRUE)
####  PR point estimation
coef<- coeftest (glm.5, vcov = sandwich)
## Coefficient
B<-coef[edadcod2 + Sexo +
        EstratoSocial +  TipodeSangreyRh +  NumeroTrabajos +
        ServiciosTrabaja+  Comorbilities+ fumar+  HorasTrabajadas, Estimate]

## coefficient Standard Error
SE<-coef[edadcod2 + Sexo +
        EstratoSocial +  TipodeSangreyRh +  NumeroTrabajos +
        ServiciosTrabaja+  Comorbilities+ fumar+  HorasTrabajadas, Std. Error]
## PR point estimation
exp(B)
## PR 95% Confidence Interval
# upper 95% CI
exp(B + qnorm(0.05 / 2) * SE)
# lower 95% CI
exp(B + qnorm(1 - 0.05 / 2) * SE)


sessionInfo( )

R version 4.0.3 (2020-10-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19042)

Matrix products: default

locale:
[1] LC_COLLATE=Spanish_Colombia.1252
[2] LC_CTYPE=Spanish_Colombia.1252
[3] LC_MONETARY=Spanish_Colombia.1252
```

2

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance
placed on this supplemental material which has been supplied by the author(s)

Occup Environ Med

[4] LC_NUMERIC=C
[5] LC_TIME=Spanish_Colombia.1252

attached base packages:
[1] readxl   Epi   foreign

Script for Bayesian correction in Stan program (INS.AJUSTE.stan):

```
data {
int y_sample;
int n_sample;
int y_sens;
int n_sens;
int y_spec;
int n_spec;
}
parameters {
real<lower=0.0, upper=1.0> p;
 real logit_Se;
  real logit_Sp;
}
transformed parameters {
real<lower=0.0, upper=1.0> Se;
 real<lower=0.0, upper=1.0> Sp;

 Se = inv_logit(logit_Se);
 Sp = inv_logit(logit_Sp);


}
model {
real p_sample = p*Se + (1-p)*(1-Sp);
y_sample ~ binomial(n_sample, p_sample);
y_sens ~ binomial(n_sens, Se);
y_spec ~ binomial(n_spec, Sp);
}
```

this R script for Bayesian analysis

```
library(cmdstanr)
library(rstan)
stanfit <- function(fit) rstan::read_stan_csv(fit$output_files())
sc_model <- stan_model("RSTAN/INS.AJUSTE.stan")

sc_data <- list(y_sample=NPositive,
            n_sample= Overall,
```

3

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

Occup Environ Med

```
                y_spec=D
                n_spec=B+D
                y_sens=A
                n_sens=A+C)


  fit1 <- sampling(sc_model, data = sc_data,
              chains= 2, iter= 20000,
              refresh=0)


  print(fit1, digits_summary = 3)


sessionInfo()
R version 4.0.3 (2020-10-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19042)


Matrix products: default


locale:
[1] LC_COLLATE=Spanish_Colombia.1252
[2] LC_CTYPE=Spanish_Colombia.1252
[3] LC_MONETARY=Spanish_Colombia.1252
[4] LC_NUMERIC=C
[5] LC_TIME=Spanish_Colombia.1252


attached base packages:
[1] stats    graphics  grDevices utils    datasets  methods
[7] base     rstan
```

References

1 Dhar M, Binu V, Mayya S. Some basic aspects of statistical methods and sample size determination in health science research. *Ayu* 2014;**35**:119. doi:10.4103/0974-8520.146202

2 MinSalud. ReTHUS. Registro Unico Nacional de Talento Humano en Salud. https://web.sispro.gov.co/THS/Cliente/ConsultasPublicas/ConsultaPublicaDeTHxIdentificacion. aspx (accessed 1 Jul 2021).

3 Altman DG, Bland JM. Statistics notes: How to randomise. *BMJ* 1999;**319**:703–4. doi:10.1136/bmj.319.7211.703

4 Lwanga SK, Lemeshow S, Organization WH. *Sample size determination in health studies : a practical manual*. World Health Organization 1991. https://apps.who.int/iris/handle/10665/40062 (accessed 1 Jul 2021).

4

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

Occup Environ Med

5   Friedman H. Simplified Determinations of Statistical Power, Magnitude of Effect and Research Sample Sizes. *Educational and Psychological Measurement* 1982;**42**:521–6. doi:10.1177/001316448204200214

6   Espelt A, Bosque-Prous M, Marí-Dell'Olmo M. Reflexiones sobre el uso de la Odds Ratio o la Razón de Prevalencias o Proporciones. *Adicciones* 2019;**31**:257. doi:10.20882/adicciones.1416

5