THE LANCET **Global Health**

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Hallal PC, Hartwig FP, Horta BL, et al. SARS-CoV-2 antibody prevalence in Brazil: results from two successive nationwide serological household surveys. *Lancet Glob Health* 2020; published online Sept 23. http://dx.doi.org/10.1016/ S2214-109X(20)30387-9.

Remarkable variability in SARS-CoV-2 antibodies across Brazilian regions: report on two successive nationwide serological household surveys

SUPPLEMENTARY MATERIALS

Antibody test

Prevalence of antibodies was assessed with a rapid point-of-care test, the WONDFO SARS-CoV-2 Antibody Test (Wondfo Biotech Co., Guangzhou, China), using finger prick blood samples. This test detects immunoglobulins of both IgG and IgM isotypes specific to SARS-CoV-2 in a lateral flow assay. Two drops of blood from a pinprick are sufficient to detect the presence of antibody. The assay reagent consists of colloidal gold particles coated with recombinant SARS-CoV-2 receptor binding domain (RBD; personal communication from the manufacturer). Following the introduction of the blood sample, reactive antibody:antigen:colloidal gold complexes, if present, are captured by antibodies against human IgM and IgG present on the "test" (T) line in the kit's window, leading to the appearance of a dark-colored line. Samples without SARS-CoV-2-reactive antibodies will not lead to appearance of this line. Valid tests are identified by a positive control line (C) in the same window. If this control line is not visible, the test is deemed non-conclusive, which is uncommon. There were only 33 non-conclusive results in the over 50,000 tests carried out in the two phases.

The rapid test underwent independent validation studies. According to the manufacturer, it has a sensitivity of 86.4% and specificity of 99.6% [\(https://en.wondfo.com.cn/product/wondfo-sars-cov-](https://en.wondfo.com.cn/product/wondfo-sars-cov-2-antibody-test-lateral-flow-method-2/)[2-antibody-test-lateral-flow-method-2/\)](https://en.wondfo.com.cn/product/wondfo-sars-cov-2-antibody-test-lateral-flow-method-2/). The tests were acquired by Brazilian Ministry of Health for population surveys and surveillance programs. A validation study carried out by the National Institute for Quality Control in Health (INCQS, Oswaldo Cruz Foundation, RJ, Brazil) showed a sensitivity of 100% and specificity of 98.7%. In an evaluation of 10 different lateral flow assays, Whitman and colleagues¹ found that the Wondfo test among the two with the best performances, with sensitivity of 81.5% and specificity of 99.1%. Our own evaluation in Brazil found a sensitivity of 77.1% and specificity of 98.0%.² By pooling the results from the four validation studies, weighted by sample sizes, sensitivity is estimated at 84.8% (95% CI 81.4%;87.8%) and specificity at 99.0% (95% CI 97.8%;99.7%).2

In early April 2020, our team conducted a household probability survey in nine cities in the state of Rio Grande do Sul,³ when the pandemic was at a very early stage in the state. Of a total sample of 4,188 subjects there were only two positive results. We believe that this survey provides a better estimate of the test's false-positive rate in the field, given that the other four studies relied on frozen samples for specificity estimation. Assuming that all cases in that survey were false positives leads to a specificity rate of 99.95%. Importantly, this conservative estimate of the specificity is higher than the estimate obtained in validation studies, further supporting that specificity in the field is higher. Whitman and colleagues, in their analyses of 10 lateral flow tests, observed "*moderate-to-strong positive bands in several pre-COVID-19 blood donor specimens, some of them positive by multiple assays, suggesting the possibility of non-specific binding of plasma proteins, non-specific antibodies, or cross-reactivity with other viruses*."1 Our findings suggest the possibility that studies using frozen serum samples may have yielded higher falsepositive rates than those associated with testing finger prick blood. We therefore used as correction parameters in the main analyses a sensitivity of 84.8% and the 99.95% specificity derived from our previous population-based survey.³ Analyses using the same sensitivity level and a specificity of 99.0% (the specificity obtained by pooling the four validation studies) are also presented below.

Calculation of wealth quintiles

Wealth was measured in this study through an asset index based on 8 asset variables. Internet at home and cable TV were yes/no questions, computer/notebook, TV, air conditioner and car were recorded as a number from zero to four or more and number of rooms and bathrooms in the house were recorded from zero to five or more. The eight variables were used in two sets of

principal components analyses, one in each survey wage. The first component accounted for 39.2% of the total variability in the first survey phase and 39.0% in the second phase. The eigenvalues were equal to 3.1 in both phases.

The scores for the first component were extracted to be used as an asset score representing wealth in the sample. The lower the score, the poorer the family and vice-versa. Mean scores by state are shown in the Figure (for the first survey phase). We see that Roraima and Pará are the states with lowest mean scores, while Santa Catarina and Rio Grande do Sul are the richest in terms of mean asset index scores.

The scores were divided into five groups, quintiles, each including one fifth of the sample, weighted by their sampling probability. The weights were calculated by dividing the population of each city (National Institute of Geography and Statistics projection for 2019) by the number of interviews

completed in the city. This is necessary to account for the large differences in population in the cities, resulting in more

Figure. Mean asset scores (Índice Econômico Nacional: IEN) by federative units in Brazil.

representative wealth groups at national level. The table below shows that the asset index was strongly associated with schooling of the head of

	ີ		ີ			
	Quintiles of the asset score					Total
Schooling	1	2	3	4	5	
1-4 years	1,594	1,020	751	548	366	4,279
5-8 years	1,388	1,035	978	745	552	4,698
9-11 years	1,555	1,702	1,887	1,966	1,646	8,756
Higher education	186	408	764	1.254	2,222	4,834
Total	4,723	4.165	4,380	4,513	4,786	22,567
Dograph chi squaro (12 df) - 1100 $n > 0$ 0001						

Table. Distribution of schooling in the sample according to the quintiles of the asset score.

Pearson chi-square (12 df) = 4100, p < 0.0001

Statistical analyses

As mentioned in the main text, the survey data was analyzed using two strategies. The first consists of accounting for the sampling design of the survey, but not for the test validity. The goal of this strategy is to provide the actual test results. In the second strategy, we accounted for both the sampling design of the survey and corrected for the test validity. The goal is to attempt to obtain estimates closer to the actual infection prevalence. Performing both strategies allows comparing results between them, such that consistency between strategies indicates that conclusions do not strongly depend on a given correction strategy. All analyses were performed using R version 3.6.1.4

To account for the sampling design, all survey data analyses were performed using the functionality provided by the "survey" package.5,6 Because municipalities were selected *a priori* and census tracts were sampled within each municipality, census tracts were treated as principal sampling units and municipalities as strata. Importantly, the fact that census tracts were sampled with probability proportionate to size and a fixed number of households was sampled in each tract, this is a self-weighted design. No weighting for population size (or any other type of weighting) was performed. Variance was estimated analytically using Taylor series linearization estimation (this procedure is described in detail elsewhere⁵).

For the second strategy, estimates of test validity are required. By pooling multiple validation studies (as described above), sensitivity was estimated to be $\hat{s} = \frac{446}{526}$ and specificity was estimated to be $\hat{e} = \frac{513}{518}$ ² However, as discussed above, the specificity is likely higher than this. Indeed, based on the first population-based survey we carried out in the state of Rio Grande do Sul in Brazil,³ a lower bound for the specificity is $\hat{e} = \frac{2}{4151}$. Therefore, unless explicitly stated otherwise, we used $\hat{s} = \frac{446}{526}$ and $\hat{e} = \frac{2}{4151}$ as the estimates of the test validity.

We obtained corrected prevalence estimates $(\hat{\theta})$ using a maximum likelihood procedure based on the following model (the rationale for this model is described elsewhere⁷):

$$
P\big(\text{observe } r \text{ positives out of } n \text{ tests} \big|\hat{\delta}\big) \\
= \binom{n}{r} \left(\hat{s}\hat{\theta} + (1-\hat{e})(1-\hat{\theta})\right)^r \left(1 - \hat{s}\hat{\theta} - (1-\hat{e})(1-\hat{\theta})\right)^{n-r},
$$

where $\hat{\delta}$ is the estimated apparent (i.e., uncorrected) prevalence.

Given $\hat{\delta}$ (obtained from the survey), \hat{s} and \hat{e} (obtained from the validation study), $\hat{\theta}$ was calculated as the value of $\vartheta_i \in \{0\%, 0.1\%, 0.2\%, ...\}$ 100%} that maximizes the likelihood based on the model above. Assuming the likelihood is unimodal, the algorithm can be implemented by first sorting the vector ϑ in ascending distance from $\hat{\delta}$, and then check if testing additional ϑ_i values result in larger likelihood values. In our analysis, this continues until the algorithm fails to identify a value that increases the maximum likelihood (among all ϑ_i values already evaluated) for 10 consecutive attempts.

To incorporate both the uncertainty of \hat{s} and \hat{e} and the sampling design in the confidence interval for $\hat{\theta}$, we used the following strategy:

- a) Design-adjusted standard errors for the unadjusted prevalence (denoted as $\sigma_{\hat{\delta}}$) were calculated by first fitting an intercept-only logistic regression model having the test result as the dependent variable. We then used this model to estimate $\sigma_{\widehat{\delta}}$ using the "predict" function.
- b) Calculate the effective sample size (N^*) i.e., the sample size that a study using simple random sampling would be expected to have so that the standard error of δ equals $\sigma_{\widehat{\delta}}$. This was calculated as $N^* = \min \left[N, \frac{\hat{\delta}(1-\hat{\delta})}{\sigma^2} \right]$ $\frac{1}{\sigma_{\delta}^2}$, where N is the actual sample size. N^{*} was rounded to the nearest integer.
- c) Generate the empirical sampling distribution of δ as $\hat{\delta}_r \sim \frac{B(N^*,\hat{\delta})}{N^*}$. Importantly, by using N^* instead of N in this step, the variance of the empirical sampling distribution is σ_{δ}^2 , thus accounting for the sampling design.
- d) Assuming that $\hat{s} \sim \frac{B(N_s,s)}{N_s}$ and $\hat{e} \sim \frac{B(N_e,e)}{N_e}$ (where $N_s = 526$ and $N_e = 4151$ denote the sample sizes used to estimate sensitivity and specificity, respectively), the empirical sampling distribution of these parameters can be obtained as $\hat{s}_r \sim \frac{B(N_s, s)}{N_s}$ e $\hat{e}_r \sim \frac{B(N_e, \hat{e})}{N_e}$, where $r \in$ ${1, ..., R}.$
- e) For each r , calculate the corresponding value of $\hat{\theta}$ (denoted by $\hat{\theta}_r$) by replacing \hat{s} , \hat{e} and $\hat{\delta}$ with \hat{s}_r , \hat{e}_r and $\hat{\delta}_r$ in the maximum likelihood estimation procedure described above. The collection of all R values of $\hat{\theta}_r$ is the empirical sampling distribution of θ estimated using parametric bootstrap.
- f) Estimate the standard error of $\hat{\theta}$ (denoted as $\sigma_{\hat{\theta}}$) as the standard deviation of the empirical distribution of θ .
- g) Update the effective sample size (N') as follows: $N' = \min \left[N^*, \frac{\theta(1-\theta)}{\sigma^2} \right]$ $\left(\frac{1-\sigma_j}{\sigma_{\hat{\theta}}^2}\right)$, rounded to the nearest integer.
- h) Calculate the effective number of positive tests as $n'_p = \delta N'.$
- i) Use n'_p and N' to calculate the exact binomial confidence interval. When n'_p is not an integer, we opted by not rounding it because, due to the small number of positive tests, any rounding would correspond to a substantial relative change in the prevalence. To overcome this issue, we calculated to confidence intervals: one for the nearest smaller integer (i.e., $\lfloor n'_p \rfloor$) and another for the nearest larger integer (i.e., $|n'_p|$). Let a_1 and b_1 respectively denote the lower and upper limits of the confidence interval using $[n'_p]$, and a_2 and b_2 denote the same for $[n'_p]$. The confidence interval for n'_p was then calculated as follows: $a = \sum_{k=1}^2 a_k w_k$ and $b =$ $\sum_{k=1}^{2} b_k w_k$, where w_k is the weight that each confidence interval receives, calculated as follows: $w_1 = 1 - (n'_p - [n'_p])$ and $w_2 = (n'_p - [n'_p])$.

For hypothesis testing under strategy 2, we compared the corrected prevalence estimates (in logit scale) among groups using Cochran's Q heterogeneity, implemented as fixed effects metaregression using the "metafor" package.⁸ For this, standard errors of logit $(\hat{\theta})$ (denoted as $\sigma_{logit(\hat{\theta})}$)

were approximated as $\sigma_{logit}(\hat{\theta}) = \frac{logit(b)-logit(a)}{2Q(0.975)}$ where $Q(.)$ is the quantile function for the Normal distribution.

In the following situations, the approach outlined in a)-i) above had to be slightly adapted:

- Uncorrected prevalence of 0%: in this situation, we calculated confidence intervals using the exact binomial method assuming simple random sampling. For hypothesis testing, groups in this situation were excluded from the comparisons.
- Uncorrected prevalence >0%, but corrected prevalence of 0%: in this situation, $l = 0\%$ and u was calculated as the $\left(1-\frac{\alpha}{2}\right)$ 100% percentile of the empirical sampling distribution of θ for a $(1 - \alpha)100\%$ confidence interval. We opted by the percentile method instead of the exact binomial method in this situation to allow for the uncertainty in $\hat{\theta}$ to be incorporated in the confidence interval. For hypothesis testing, groups in this situation were excluded from the comparisons.

References

- 1. Whitman JD, Hiatt J, Mowery CT, et al. Test performance evaluation of SARS-CoV-2 serological assays. *(unpublished*) 2020.
- 2. Pellanda LC, Wendland EM, McBride AJA, et al. Sensitivity and specificity of a rapid test for assessment of exposure to SARS-CoV-2 in a community-based setting in Brazil. Preprint at <https://www.medrxiv.org/content/10.1101/2020.05.06.20093476v1> (2020).
- 3. Silveira MA, Barros AJD, Horta BL, et al. Repeated population-based surveys of antibodies against SARS-CoV-2 in Southern Brazil. *Nat Med* 2020; doi: 10.1038/s41591-020-0992-3.
- 4. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. URL[: https://www.R-project.org/](https://www.r-project.org/) (2018).
- 5. Lumley, T. Analysis of complex survey samples. *J Stat Soft* 2004; **9**: 1-19.
- 6. Lumley, T. survey: analysis of complex survey samples. R package version 3.35-1. URL: [https://CRAN.R-project.org/package=survey](https://cran.r-project.org/package=survey) (2019).
- 7. Lewis FI, Torgerson PR. A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic. *Emerg Themes Epidemiol* 2012; **9:** 9.
- 8. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J Stat Soft* 2010; **36**: 1-48.

Supplementary Figure 1. Sample selection flowchart for the two phases of the study.

Supplementary Table 1. Characteristics of the 83 municipalities with 200 or more tests in both phases, other municipalities included in the study and the remaining Brazilian municipalities. Reported cases and deaths refer to cumulative numbers up to May 23.

(*) Source[: https://covid.saude.gov.br/](https://covid.saude.gov.br/)

There are 5,565 municipalities in Brazil. We compared population sizes, reported COVID-19 cases and deaths and the Human Development Index 4 in three groups of cities: the 83 where it was possible to conduct 200 or more tests during both survey waves, the remaining 50 cities included in the original sample, and the other 5,432 cities in the country (Supplementary Table 1). Cities with 200 or more tests tended to have larger populations and higher rates of reported cases and deaths than those with fewer than 200 tests, or the remaining cities in the country. The Human Development Index of the first two groups tended to be higher than in the third group of cities. Data on the index were obtained from:

PNUD. Atlas do Desenvolvimento Humano no Brasil. Brasilia: Programa das Nacoes Unidas para o Desenvolvimento; 2010.

Supplementary Table 2. Sample distribution according to sociodemographic characteristics.

Supplementary Table 3. Results from the two survey phases in the 133 cities. Corrected for sample design and test parameters (sensitivity 84.8% and specificity 99.95%).

Supplementary Figure 2. Location of the 13 cities in the Amazon region with the highest prevalence in the study. Satellite images are from Google Earth.

Supplementary Table 4. Results from the two survey phases in the 133 cities. Corrected for sample design and test parameters (sensitivity 84.8% and specificity 99.0%).

Supplementary Figure 3. Time trends in reported deaths between the start of the epidemic in each region and May 13, 2020 (source: https://covid.saude.gov.br)

Supplementary table 5. Odds ratios for antibody prevalence according to skin color.