

# Miller School Professors Develop Model To Correct COVID-19 Sampling Bias

COVID-19 testing has been a standard approach to estimate the prevalence of disease and assist with public health decision-making to mitigate the spread of the virus. But the sampling methods used are often biased – such as the over-representation of symptomatic people who are more likely to be tested. This leads to a biased sample that does not reflect a population's true COVID-19 infection prevalence.

A new [study](#) published in the *Journal of Theoretical Biology* presents a simple way to correct COVID-19 testing bias. J. Sunil Rao, Ph.D., professor and director of the Division of Biostatistics, and Daniel A. Díaz-Pachón, research assistant professor – both with the Department of Public Health Sciences at the University of Miami Miller School of Medicine – have developed a bias correction methodology to use on already collected information that reduces estimation errors dramatically.



“The derived correction is simple and can be easily implemented in studies across the world,” said Dr. Díaz-Pachón, lead author of the study. “We want the model to help guide public health decisions.”

The researchers noted sample bias also occurs in meta-analyses due to publication bias – papers that show no treatment effect (null hypothesis) are less likely to be published, which leads to overestimates of the effect of treatment. They discovered that modeling the null-favoring censoring mechanism could be used to correct COVID-19 sampling bias.

“In the case of COVID-19 testing studies, biased sampling occurs when patients who are symptomatic are more likely to be tested than those who are asymptomatic,” said Dr. Rao, senior author of the study. “This essentially represents convenience samples. These studies are therefore missing key information about how specific populations are being affected by COVID-19. Random sampling is costly and inefficient and thus not widely done.”

To provide a solution, Dr. Rao and Dr. Díaz-Pachón developed a model that led to a three-step correction general enough to allow customization. For instance, symptomatology may be extended to reflect different subpopulations, such as racial and ethnic subgroups, age groups, risk groups, and varying environments. The model can also be generalized to index more than one type of test.

“Once we formalized our mathematical model and evaluated the nature of the bias in prevalence of COVID-19 convenience samples, we realized that this bias was modifying the true prevalence, multiplying it by a value coming from the population,” said Dr. Díaz-Pachón. “Therefore, in order to correct it, we needed to divide the biased estimation by the sample estimate of such value.”

To do so, Dr. Díaz-Pachón and Dr. Rao extracted as much information as they could from the sample, and when it was not possible to advance any more, they resorted to a concept of information theory called maximum entropy, which reduces bias as much as possible, given their current level of knowledge of the sample.

“The math was consistent and the model consonant with observations in that when it was compared to real data we could see that it resembled reality much better than the original biased estimate,” said Dr. Díaz-Pachón. “However, the model provides a bias correction, not an elimination of bias.”

The study provides examples of two real-life scenarios – COVID-19 outbreaks on a cruise ship and in Lombardy, Italy – where the model is implemented. The results show the proposed method achieves significant reductions in sampling bias, particularly for the overall prevalence.

Since the publishing of the paper, which was also included in [the World Health Organization COVID-19 research repository](#), Dr. Rao and Dr. Díaz-Pachón have begun new collaborations in Israel and Colombia. In Israel, they are working with Central Statistics Bureau scientific leaders who are conducting their National Serological Survey for COVID-19. In Colombia, they

are collaborating with researchers at the State of Antioquia's Public Health Laboratory who are doing various novel surveillance testing studies for COVID-19.

Content Type Article