

# Do Face Masks Reduce COVID-19 Spread in Bangladesh? Are the Abaluck *et al.* Results Reliable?

Denis G. Rancourt, PhD

Researcher, Ontario Civil Liberties Association ([ocla.ca](http://ocla.ca))

This article is published here:

<https://denisrancourt.ca/entries.php?id=106>

20 September 2021

## Table of Contents

Purpose

Summary

Can the chosen antibody test be used in this application?

*Is the antibody assay specific for SARS-CoV-2?*

*Is the antibody assay validated for use in Bangladesh?*

*Was “spectrum bias” duly examined by InBios and Abaluck et al.? Are the positives reliable?*

*Conclusion regarding the serology test*

Are the control and treatment arms valid (comparable)?

*Treatment alone versus adding super-treatment interventions*

*Science of the stress-immune relationship*

*Mechanisms of bias from the super-treatment interventions*

Is the size of the trial sufficient for the results to be reliable?

*All adults, 18 through 60+ years old, both mask types together*

*Oldest age group, 60+ years old, surgical masks only*

Conclusion

References

Appendix A: Media reviews of the Abaluck *et al.* (2021) mask study

My competence to review science about COVID-19

## Purpose

**“This really should be the end of the debate,”** says Ashley Styczynski, an infectious-disease researcher at Stanford University in California and a co-author of the preprint describing the trial. The research **“takes things a step further in terms of scientific rigour”**, says Deepak Bhatt, a medical researcher at Harvard Medical School in Boston, Massachusetts, who has published research on masking. — *Nature* | News | 09 September 2021 | [“Face masks for COVID pass their largest test yet”](#)

The leading trend-setting mainstream media and institutional public relations offices have been unreservedly enthusiastic about “the Bangladesh mask study” (see **Appendix A**).

Here, I review the methods and results of that study by Abaluck *et al.* (2021) published as a working paper by Innovations for Poverty Action (IPA): [“The Impact of Community Masking on COVID-19: A Cluster-Randomized Trial in Bangladesh”](#), 01 September 2021.

The study’s stated primary outcome regarding the benefits of face masks is “symptomatic SARS-CoV-2 seroprevalence”, meaning the prevalence during the study period of individuals self-reporting COVID-like symptoms who also test positive using a laboratory blood test presumed to be specific for SARS-CoV-2.

## Summary

The cluster-randomized trial study of Abaluck *et al.* (2021) is fatally flawed, and therefore of no value for informing public health policy, for two main reasons:

- I. The antibody detection was performed using a single commercial FDA emergency-use-authorized (EUA) serology test that is not suitable for the intended application to SARS-CoV-2 in Bangladesh (not calibrated or validated for populations in Bangladesh; undetermined cross-reactivity against broad-array IgM antibodies, malaria, influenza, etc.).
- II. The participants (individual level, family level, village level) in the control and treatment arms were systematically handled in palpably different ways that are

linked to factors established to be strongly associated to infection and severity with viral respiratory diseases, in particular, and to individual health in general.

These disjunctive fatal flaws are explained below. Either one is sufficient to invalidate the results and conclusions of Abaluck *et al.*

Furthermore, the Abaluck *et al.* symptomatic seroprevalence (SSP) results are *prima facie* statistically untenable. The treatment-to-control differences in numbers of symptomatic seropositive individuals are too small to rule out large unknown co-factor, baseline heterogeneity, and study-design bias effects. In addition, they are at best borderline significant, in terms of purely ideal-statistical estimations of uncertainty. Finally, the practice of using whole households while reporting on an individual basis, introduces unknown correlations/ clustering, and vitiates the mathematic assumptions that underlie the statistical method.

## **Can the chosen antibody test be used in this application?**

### *Is the antibody assay specific for SARS-CoV-2?*

A single laboratory test was used in the Abaluck *et al.* (2021) study: the “SCoV-2 Detect™ IgG ELISA” test kit (InBios, Seattle, Washington).

Here, ELISA stands for enzyme-linked immunosorbent assay, which is one of three main assay methods for routinely detecting or quantifying antibodies. IgG is a class of immunoglobulins. For the non-expert, two of the five classes of immunoglobulins, which are of relevance in the present critique, can be described as follows:

- Immunoglobulin M (IgM) – IgM antibodies are produced as a body's first response to a new infection or to a new "non-self" antigen, providing short-term protection. They increase for several weeks and then decline as IgG production begins.
- Immunoglobulin G (IgG) – About 70-80% of the immunoglobulins in the blood are IgG. Specific IgG antibodies are produced during an initial infection or other antigen exposure, rising a few weeks after it begins, then decreasing and stabilizing. The body retains a catalog of IgG antibodies that can be rapidly reproduced whenever exposed to the

same antigen. IgG antibodies form the basis of long-term protection against microorganisms. In those with a normal immune system, sufficient IgG is produced to prevent re-infection. Vaccinations use this process to prevent initial infections and add to the catalog of IgG antibodies, by exposing a person to a weakened, live microorganism or to an antigen that stimulates recognition of the microorganism. — *Merk Manuals* | [Immunoglobulins \(IgA, IgG, IgM\)](#) | accessed on 15 September 2021

Abaluck *et al.* (2021) state “This assay detects IgG antibodies against the spike protein subunit (S1) of SARS-CoV-2.” This statement is incorrect.

None of the official documents about the assay claim that the assay detects “the spike protein subunit (S1) of SARS-CoV-2”, or any part(s) of the spike protein. Rather, only a broad claim is ever made, of the type “The SCoV-2 Detect IgG ELISA is authorized for the detection of antibodies to SARS-CoV-2 in human serum or plasma” or “INTENDED USE: The SCoV-2 Detect™ IgG ELISA is an in vitro diagnostic test for the qualitative detection of IgG antibodies to SARS-CoV-2 in human serum or plasma”:

- [IFU LBL-0113-03 \(English\)](#) (Instructions for Use) (19 May 2021)
- [Brochure COVE-G](#)
- [FDA EUA Letter of Authorization COVE-G](#) (14 May 2021)
- [Health Care Provider Fact Sheet COVE-G](#) (14 May 2021)

These documents are also available on the FDA website:

<https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/in-vitro-diagnostics-euas>

The only mention of “spike”, which I could find, is that the FDA webpage “[EUA Authorized Serology Test Performance](#)” (“Content current as of 18 August 2021”, accessed on 14 September 2021) has the title of the section for this assay as:

### **InBios SCoV-2 Detect IgG ELISA**

**Developer:** InBios

**Test:** SCoV-2 Detect IgG ELISA

**Technology:** ELISA

**Target:** Spike

The latter FDA (Test Performance, 2021) webpage provides the independent scientific assessment in the “Test Facts” that were used for FDA EUA approval as [“NCI's Frederick National Laboratory for Cancer Research Evaluation Report”](#) (dated 13 July 2021; accessed on 14 September 2021).

The said independent scientific assessment (FNLCR, 2021) is the reference document for evaluating the assay used by Abaluck *et al.* (2021). The FNLCR (2021) report makes it clear that not only was the assay not validated for detecting any specific SARS-CoV-2 IgG antibody, but it was also not validated for any ability to distinguish IgM and IgG:

“The positive samples selected may not reflect the distribution of antibody levels in patient populations that would be evaluated by such a test. Because all samples are positive for both IgM and IgG, this evaluation cannot verify that tests intended to detect IgM and IgG antibodies separately detect these antibodies independently.”

Given the nonspecificity of IgM — by its very nature as an initial broad-array immune response — this means that the assay may have a high potential for cross-reactivity with a large spectrum of infections or conditions.

The manufacturer of the assay (InBios) reports having made an in-house (not independent) evaluation of “Cross-Reactivity (Analytical Specificity)” and reports no cross-reactivity for several antibodies to other viral infections and autoantibodies, based on small numbers ( $n = 3-8$ ) of unspecified reference samples, as (InBios, IFU LBL-0113-03, 2021):

Category	Number of samples tested	Number reactive
Anti-Influenza A/B	7	0
Anti-Hepatitis B	5	0
Anti-Hepatitis C	5	0
Anti-Nuclear Antibody	5	0
Rheumatoid Factor	18	3
Human Anti-Mouse Antibody	3	0
Anti-HIV	8	0
Anti-Respiratory Syncytial Virus	4	0
Normal Human Sera	180	0

Presumably, the reference samples were chosen to have specific IgG of the tested viral infections, and would therefore have little or no residual IgM initially induced by the tested infections, since IgG is generated as IgM decreases as functions of time from onset of symptoms.

From this Table (InBios, IFU LBL-0113-03, 2021), one might ask: Since cross-reactivity for rheumatoid factor was detected (3/18) by testing 18 samples, why were more samples not used for the other diseases (at least 18 samples, say)? After all, there is no lack of influenza standards, for example. Otherwise, with the small number of samples used, it is entirely possible to have missed large incidences of cross-reactivity.

As it stands, cross-reactivity is reported solely for “Rheumatoid Factor” (3/18) (InBios, IFU LBL-0113-03, 2021). Given this known cross-reactivity of the assay, Abaluck *et al.* should have obtained baseline prevalence of rheumatoid arthritis and Sjogren's syndrome in their control and intervention arms, especially for their most elderly cohorts (50-60 and 60+ years) and for the two types of face masks, or they should have ruled out these conditions in their elderly “symptomatic seropositive” individuals, especially in view of their most surprising results (their Figure 3). Abaluck *et al.* did not do this (did not report doing this).

Yadouleton *et al.* (2021) studied cross-reactivity (specificity) of the InBios SCoV-2 Detect™ IgG ELISA assay, and of another ELISA assay nominally for SARS-CoV-2 antibodies. Of 60 pre-COVID (2019) samples from Benin, they found that the InBios assay gave many samples that were near the positive/negative threshold (“cut-off”) (their Figure 1A, fourth panel). They concluded, from the results for both assays: “acute malaria is the most plausible explanation for unspecific SARS-CoV-2 ELISA reactivity in pre-pandemic controls”, and found false positive rates as high as 25% (for the non-InBios assay).

The study of Yadouleton *et al.* (2021) is especially relevant because “Bangladesh is one of the four major malaria-endemic countries in South-East Asia having approximately 34% of its population at risk of malaria [...] with a prevalence ranging between 3.1% and 36%” (Islam *et al.*, 2013). Abaluck *et al.* did not report having surveyed or screened for past or present infections of malaria among their study subjects.

### *Is the antibody assay validated for use in Bangladesh?*

The short answer is “no”. The long answer is as follows.

To start, we need accurate definitions of test specificity and sensitivity, which are provided, in the words of the FDA (Test Performance, 2021), as:

The performance of these [EUA authorized serology] tests is described by their "sensitivity," or their ability to identify those with antibodies to SARS-CoV-2 (true positive rate), and their "specificity," or their ability to identify those without antibodies to SARS-CoV-2 (true negative rate).

There are two major problems with application of the InBios antibody assay to populations in Bangladesh.

The first major problem is that the performance of the emergency utilization authorized InBios test has never been evaluated for a real-world population; not in the USA, and not in Bangladesh. In the words of the independent evaluators (FNLCR, 2021) (p. 4):

Samples used in this evaluation were not randomly selected, and sensitivity (PPA) and specificity (NPA) estimates in this report may not be indicative of the real-world performance of the InBios International Inc. SCoV-2 Detect™ IgG ELISA. [...]

### **1.3 Important caveats**

Sensitivity and specificity estimates in this report may not be indicative of the real world performance of the InBios International Inc. SCoV-2 Detect™ IgG ELISA. [...]

The number of samples in the panel is a minimally viable sample size that still provides reasonable estimates and confidence intervals for test performance, and the samples used may not be representative of the antibody profile observed in patient populations.

The second major problem is as follows.

The InBios test is based on optical density (OD) measurements through the ELISA solution in the final step of the assay: the more reactive the sample (to the ELISA substrate intended to bind the target antibody), the greater the OD. The measured OD is divided by "the average OD plus three standard deviations" for many reference samples presumed to be free of the target antibody. This ratio ( $OD_{\text{sample}}/OD_{\text{cut-off}}$ ), called the "Immunological Status Ratio" (ISR), is used to discriminate "positive" ( $ISR \geq 1.1$ ) and "negative" ( $ISR \leq 0.9$ ) samples. The manufacturer considers ISR values of  $>0.9$  through  $>1.1$  to be "borderline"/undetermined results.

In the words of the manufacturer (InBios, IFU LBL-0113-03, 2021) (p. 10):

The assay cut-off value was determined by screening a large number (>100) of normal human serum (NHS) samples that were collected [in the USA] prior to the COVID-19 outbreak (~November, 2019). The cut-off selection was performed by estimating the mean of the negative specimens plus three (3) standard deviations.

Therefore, the determination of  $OD_{\text{cut-off}}$  is critical and its value depends on the population from which one draws the so-called NHS samples. We can presume that InBios drew its NHS samples from a USA population, and that its arbitrary choices of “1.1/0.9 ISR thresholds” and “plus three (3) standard deviations” were made in order to “make it work”. That is, in order to resolve “positive” from “negative” serum samples, from USA residents known independently to test positive for SARS-CoV-2.

It is not reasonable to expect that the thus adopted test values ( $OD_{\text{cut-off}}$ , and 1.1/0.9 ISR thresholds) determined using “NHS” from USA residents would apply to a population of Bangladesh citizens, because the pre-COVID “normal human serums” from Bangladesh citizens would be significantly different, regarding the prevalence of antibodies to various viral infections, autoantibodies, and cross-reactivity with immune-response products from various other infections (e.g., malaria) and conditions (e.g., rheumatoid arthritis, Sjogren's syndrome).

Indeed, even entirely within the USA, Kaufman *et al.* (2021), in their large study of “More than 2.4 million SARS-CoV-2 IgG serology (initiated April 21, 2020) and 6.6 million nucleic acid amplification testing (NAAT) (initiated March 9, 2020) results on persons from across the United States as of July 10, 2020”, found that: “SARS-CoV-2 IgG positivity was observed in 91% (19,434/21,452) of individuals tested after a positive [nucleic acid amplification testing] NAAT result and in 10% (7,831/80,968) after a negative NAAT result. Factors associated with seropositivity include age, region of patient residence, and interval between NAAT and IgG serology.”

To be clear, Kaufman *et al.* (2021) found that both the rate of IgG positivity among NAAT-positive individuals (~sensitivity) and the rate at which NAAT-negative individuals had subsequent IgG positivity (~false-positive rate) differed significantly with respect to geographic area within the USA: 93.4% to 86.2% and 16.4% to 4.8%, respectively, in going from the 5-state NE area (NY/NJ/MA/RI/CT) to all other states (their Figure 3).

Therefore, we must assume that there can be a large systematic difference in serology test performance and/or in population immunological response or characteristics in going from the USA to Bangladesh. The estimated magnitude of this systematic effect, indicated by the extensive results of Kaufman *et al.* (2021) for different geographical regions in the USA, is large



enough to invalidate those results from Abaluck *et al.* that involve small differences in numbers of tested individuals, such as the impact of surgical masks on the most elderly cohorts, even if there were not the serious validation problems outlined above for the InBios test.

Furthermore, purely in terms of population immunology, do USA and Bangladesh populations have different prevalences, at any given time, of broad-array IgM, which the InBios test is not established to resolve from IgG?

Specifically, the spectrum of disease prevalence in Bangladesh is dramatically different than in the USA. Bangladesh has a “high” degree of risk (2020) for ([The World Factbook](#)): bacterial and protozoal diarrhea, hepatitis A and E, typhoid fever, dengue fever, malaria, leptospirosis, and rabies; and an obesity rate of 3.6 % (2016), compared to the USA obesity rate of 36.2% (2016) (adult prevalence rate).

Serum matrix effects (“cross-reactivity”) must be expected to be large and different for Bangladesh, compared to the USA. Irrespective of anything else, or of any manufacturer’s claims, Abaluck *et al.* (2021) should have stringently tested a representative array of known (independently and reliably determined) positive and negative serum samples from Bangladesh, using the InBios test as provided. Without this minimal precaution of upfront verification to rule out differences and to validate test utility, their test results are useless for the intended scientific purposes.

*Was “spectrum bias” duly examined by InBios and Abaluck et al.? Are the positives reliable?*

The answer is “no”, at least on the basis of what is reported.

“Spectrum bias” is the unavoidable variation of performance of a test arising from the frequency distribution (“spectrum”) of values that are being measured by the test in the given tested population (for example, see: Usher-Smith *et al.*, 2016).

Two problems occur.

- (1) At calibration: a test can have a significantly different actual performance than the performance evaluated using any set or array of known samples if the manufacturer’s

calibration (for setting of cut-off and undetermined range, and for assay protocol development) uses solely means and standard deviations, without regard to the shape of the distribution of test measurements (OD values) of the calibration samples (the “>100 of normal human serum (NHS)” samples used by InBios). This can produce misleading and over-enthusiastic test performance characteristics, and it again demonstrates the importance of using representative calibration samples.

- (2) In the field: a test can have significantly different performances (sensitivity, specificity) on different populations having different distributions of test measurements (OD values), even if the populations are otherwise comparable (comparable cross-reactive pathogens, co-factors, age structure, health status, etc.).

One simple consequence of the “spectrum bias” effect is that, in populations with low prevalence, many of the test results are close to the positive/negative threshold value, leading to particularly large errors, in general. This is why the FDA states (FDA, Test Performance, 2021) (p. 2):

In low prevalence populations, the result of a single antibody test is not likely to be sufficiently accurate to make an informed decision regarding whether or not an individual has had a prior infection or truly has antibodies to the virus. A second test, typically one assessing for the presence of antibodies to a different viral protein, generally would be needed to increase the accuracy of the overall testing results.

This is also why the FDA (Test Performance, 2021) (p. 47) estimates a theoretical 95% confidence interval of (50.5%, 100%) in the positive predictive value (PPV) (probability of a positive being correct) for 5% population prevalence for the InBios test, despite the stellar EUA evaluation numbers.

This means that, depending on “prevalence” of the assay-reactive condition in the Bangladesh study populations of Abaluck *et al.*, the reliability of a positive determination can be 50% or less for small prevalence. Abaluck *et al.* report symptomatic prevalences of 0.76% (control arm) and 0.68% (intervention arm).

In the present case, the “test measurement” or “value that is being measured” is the above-described ratio ( $OD_{\text{sample}}/OD_{\text{cut-off}}$ ), called the “Immunological Status Ratio” (ISR), obtained for a given serum sample using the InBios assay. It is a continuous variable, and it is obviously prone to “spectrum bias” since the manufacturer even defines an undetermined region, for  $ISR > 0.9$  through  $> 1.1$ , rather than simply a definite positive/negative threshold value.

Therefore, if InBios wanted users and evaluators to gauge the potential for “spectrum bias”, then it would, among other things, publish the distribution of ISR values of its large number of so-called normal human serum (NHS) samples that were collected in the USA prior to COVID (InBios, IFU LBL-0113-03, 2021). I could not find such information, or any discussion of this issue. Likewise, the FNLCR (2021), in its evaluation of the test, discloses only positive/negative status, not ISR values for the evaluation samples.

Similarly, Abaluck *et al.* do not disclose their ISR values, do not show distributions of ISR values, and do not even state how many of their samples gave “undetermined” (“equivocal”) ISR values on initial measurement (Abaluck *et al.*, 2021):

[...] the immunological status ratio (ISR) was calculated as the ratio of optical density divided by the cut-off value. Samples were considered positive if the ISR value was determined to be at least 1.1. Samples with an ISR value 0.9 or below were considered negative. Samples with equivocal ISR values were retested in duplicate, and resulting ISR values were averaged.

For example, are the distributions of ISR values different for the control and intervention arms? We do not know.

### *Conclusion regarding the serology test*

In conclusion, the FDA emergency-use-approved (EUA) InBios serology test was improperly applied by Abaluck *et al.* (2021):

- i. It is not specific to SARS-CoV-2, since it has undetermined cross-reactivity against broad-array IgM antibodies (n=0), undetermined cross-reactivity with other corona viruses (n=0), probable cross-reactivity with malaria (peer-reviewed article), known cross-reactivity with rheumatoid factor (n=18), insufficiently tested cross-reactivity with influenza A/B (n=7), hepatitis B (n=5), hepatitis C (n=5), respiratory syncytial virus (n=4), and others, undetermined cross-reactivity (n=0) with the high-risk pathogens endemic to Bangladesh (bacterial and protozoal diarrhea, hepatitis A and E, typhoid fever, dengue fever, malaria, leptospirosis, and rabies), and unknown comparative serum matrix effects in USA and Bangladesh.

- ii. It has not been validated with any actual population, whether in the USA or Bangladesh, and is calibrated solely using USA serum samples.
- iii. It is not calibrated or validated for Bangladesh, and cannot be used as-given on residents of Bangladesh.

I find it unacceptable that a test that is not approved for patients —

LIMITATIONS: ... • Assay results should be interpreted only in the context of other laboratory findings and the total clinical status of the patient. (InBios, IFU LBL-0113-03, 2021) (p. 12)

— would be used to diagnose participants in a trial, as having COVID-19, without any clinical evaluation beyond self-reporting of symptoms with survey questions, in order to justify long-term application of a treatment to millions of people, which has known and unknown associated harms (Rancourt. 2021).

## **Are the control and treatment arms valid (comparable)?**

Let me start by stating the obvious, since it seems to have escaped detection by virtually all media and public-relations reviewers (including the folks at *Nature*): A trial in which the researchers spend significant resources to convince the non-control group to accept or adopt the treatment is not a “randomized” trial, nor is it “controlled”. Rather, it is a trial in which one group is chosen to be intrusively manipulated to receive the treatment, whereas the other group is free from this manipulation. The trial design is not one in which the treatment and control groups are distinguished by the presence or absence of treatment, as the sole systematic difference. In addition, in this case, individuals in both groups are free to adopt the treatment or not, and that choice is anything but random, in both groups. If anything, the study of Abaluck *et al.* is in-effect merely another comparative study, but with extensive researcher interference.

### *Treatment alone versus adding super-treatment interventions*

The study of Abaluck *et al.* (2021) suffers from a major difficulty: the researchers must apply significant and repeated interventions (in a campaign to induce acceptance of the treatment of mask wearing) to the treatment arm, while preventing those interventions in the treatment arm from inducing bias in the outcome.

In other words, the cluster-randomized study is worse than merely unblinded. It is a case in which the treated individuals are not solely subjected to the treatment (mask wearing), but are additionally subjected to the sustained and multi-faceted campaign of interventions to induce acceptance of the treatment.

It is one thing to design and evaluate interventions intended to generate mask use, but it is quite another thing to measure the health impact of increased mask use alone, without introducing co-factors arising from the interventions.

One way to reduce potential bias would have been to measure prevalence of the disease solely in families in the treatment arm (treatment villages) randomly selected not to be subjected to the interventions, if that were possible with redesigned interventions. However, this was not done. Prevalence in the treatment arm was measured in the same individuals and families that were subjected to the interventions.

This is not a fatal flaw if there are compelling and empirically supported reasons to believe that the additional (super-treatment) measures cannot affect the outcome. However, in this case, the opposite is true: there are compelling reasons to expect that the super-treatment measures affect the outcome, as explained below.

The basic super-treatment intervention consisted of the following elements, as described by Abaluck *et al.* (2021):

To emphasize the importance of mask-wearing, we prepared a brief video of notable public figures discussing why, how, and when to wear a mask. The video was shown to each household during the mask distribution visit and featured the Honorable Prime Minister of Bangladesh Sheikh Hasina, the head of the Imam Training Academy, and the national cricket star Shakib Al Hasan. During the distribution visit, households also received a brochure based on WHO materials depicting proper mask-wearing.

We implemented a basic set of interventions in all treatment villages, and cross-randomize additional intervention elements in randomly chosen subsets of treatment villages to investigate whether those have any additional impact on mask-wearing. The basic intervention package consists of five main elements:

1. One-time mask distribution and promotion at households.
2. Mask distribution in markets on 3-6 days per week.
3. Mask distribution at mosques on three Fridays during the first four weeks of the intervention.
4. Mask promotion in public spaces and markets where non-mask wearers were encouraged to wear masks (weekly or biweekly).
5. Role-modeling and advocacy by local leaders, including imams discussing the importance of mask-wearing at Friday prayers using a scripted speech provided by the research team.

Participants, mask promoters, and mask surveillance staff were not blinded as intervention materials were clearly visible.

### *Science of the stress-immune relationship*

The science background to understand why the interventions of Abaluck *et al.* would have an impact on prevalence is as follows.

First, researchers performing comparative trials for outcomes involving immune response must make themselves aware that ordinary psychological stress significantly impacts immune response, and that psychoneuroimmunology is a large field of research (Ader and Cohen, 1993).

Social status, within a specific dominance hierarchy, is a major predictor of chronic stress, in social animals including humans (Cohen *et al.*, 1997a) (Sapolsky, 2005), which, in turn, may be the dominant determinant of individual health, disease burden, and longevity (Cohen *et al.*, 2007).

Ordinary psychological stress is known to be a dominant factor in making an individual susceptible to viral respiratory disease symptomatic infection, and to increase the severity of

the infection (Cohen et al., 1991). Also, social isolation (paucity of social-network interactions), in addition to individual psychological stress, is known to have an added impact on the individual's susceptibility to viral respiratory disease (Cohen et al., 1997b).

Furthermore, there is a large age gradient: extended periods of psychological stress are known to have more deleterious health effects in elderly persons than in younger persons (Prenderville et al., 2015).

The stress-immune relationship, however, is not simply a monotonic function of integrated intensity. Frequency and duration are pivotal: chronic or long-term stress harms immune response, whereas short-term adaptive stress enhances immune response. The often-cited review by Dhabhar (2014) has:

Short-term (i.e., lasting for minutes to hours) stress experienced during immune activation enhances innate/primary and adaptive/secondary immune responses. Mechanisms of immuno-enhancement include changes in dendritic cell, neutrophil, macrophage, and lymphocyte trafficking, maturation, and function as well as local and systemic production of cytokines. In contrast, long-term stress suppresses or dysregulates innate and adaptive immune responses by altering the Type 1–Type 2 cytokine balance, inducing low-grade chronic inflammation, and suppressing numbers, trafficking, and function of immunoprotective cells.

Peters *et al.* (2021) have reviewed these concepts and the known science for the relevance to COVID-19. They pointed out that “the socioeconomic issues and various aspects of the Western type lifestyle that are closely associated with psychosocial stress have recently been reported to contribute to COVID-19”. Their ultimate aim is to “clarify whether psychosocial interventions have the potential to optimize neuroendocrine-immune responses against respiratory viral infections during and beyond the COVID-19 pandemic.”

### *Mechanisms of bias from the super-treatment interventions*

Given the above-reviewed knowledge, it seems clear to me that Abaluck *et al.* (2021) have failed to consider a critical issue in their study design. Their interventions are interpersonal and societal interactions. All such interactions either induce or relieve psychological stress experienced by the individual, to different degrees and of different durations.

Specific elements (1 to 5) of the “basic intervention package” implemented by Abaluck *et al.* can be anticipated to modulate psychological stress in the following ways:

- (1) The distribution visit to each household in the treatment arm: “The video was shown to each household during the mask distribution visit and featured the Honorable Prime Minister of Bangladesh Sheikh Hasina, the head of the Imam Training Academy, and the national cricket star Shakib Al Hasan. During the distribution visit, households also received a brochure based on WHO materials depicting proper mask-wearing.”

→ Such a visit would provide (as it appears to have been intended to provide) hierarchical validation to the family members, thus raising the experienced social status, and reducing the dominance-hierarchy stress, experienced by lower strata, below its pre-visit long-term baseline value.

- (2, 3) The masks themselves would serve as a visual symbol of belonging to this thereby privileged group, and the regular mask distributions (in markets and at mosques) would be a constant interactive confirmation of an appreciative and caring hierarchical authority; all of which boosts the perceived increased social status, and reduces or displaces dominance-hierarchy stress.

- (4) “Mask promotion in public spaces and markets where non-mask wearers were encouraged to wear masks (weekly or biweekly)”: “mask promoters patrolled public areas a few times a week and asked those not wearing masks to put on a mask.” (Abaluck *et al.* found that excluding this element produced an increase in mask use of 10.9%, compared to 28.4% when it was included.)

→ Such interactions are classic short-term, mostly unpredictable and repeated stress events, precisely of the type that “enhances innate/primary and adaptive/secondary immune responses” (Dhabhar, 2014).

- (5) “Role-modeling and advocacy by local leaders, including imams discussing the importance of mask-wearing at Friday prayers using a scripted speech provided by the research team”

→ “Role-modeling” would again strengthen the perceived increased social status, and reduce dominance-hierarchy stress. “Advocacy” can be oppressive, but it can also be of a more collaborative nature, which would work better when the advocate cannot surveil or enforce, and which would again work to reduce long-term dominance-hierarchy stress below the pre-study baseline.



Therefore, given what is known about stress-immune relations, the super-treatment interventions applied by Abaluck *et al.* would thereby enhance immune responses in the participants in the treatment arm, and consequently would reduce the probability of developing symptoms and of being infected, irrespective of any effect arising from filtration by the face masks.

Peters *et al.* (2021) envisage and argue for preventative treatment by stress management strategies precisely for COVID-19.

Furthermore, a successful socializing and educational campaign to the effect that face masks provide safety would be anticipated to create a bias towards a smaller tendency to recognize and report symptoms. In the Abaluck *et al.* study, symptoms were reported by phone or in person survey-interviews with the heads of families.

Thus, the trial design in the Abaluck *et al.* study has foreseeable built-in biases probably acting in the same direction. Their experimental design with interventions is fatally flawed, and the results are therefore of no value, irrespective of the problems with the blood test.

## **Is the size of the trial sufficient for the results to be reliable?**

### *All adults, 18 through 60+ years old, both mask types together*

There were approximately 170 K individuals in each arm of the study, which is a large number (Abaluck *et al.*, 2021). This does not in itself guarantee statistically reliable results, depending on the sizes of the cohort-specific treatment-to-control differences being reported, compared to the relevant theoretical standard deviations of the presumed purely ideal-statistical variations.

(I emphasize “ideal-statistical” because, as explained below, Abaluck *et al.* used households of closely interacting family members but then reported individual-based results, which vitiates the underlying theoretical assumptions of “independent, uncorrelated and random” in all the (ideal) statistical calculations of uncertainties and confidence intervals.)

From this sample size (170 K), there were approximately 13.5 K individuals in each arm who were reported to have developed “COVID-like symptoms” within the measurement time of the study: 13,273 (7.62%) (treatment), 13,893 (8.62%) (control). The control-treatment difference of 620, is significant since it is 5 times greater than the ideal-statistical standard deviations of the numbers prior to taking their difference,  $\sqrt{13.5 \text{ K}}$ .

The numbers of symptomatic individuals having positive serology test results, and their treatment-control differences, however, are much smaller. Abaluck *et al.* (2021) chose not to report these numbers but instead reported only “symptomatic seroprevalence” (SSP), as percentages, after accounting for the rates (~40 %) of consent to the blood test (RCB): 0.68 % (treatment), 0.76 % (control).

I work backwards from their numbers to calculate the numbers of symptomatic individuals having positive blood test results, as follows:

**Treatment arm:**

178,288 participants x 0.0068 (SSP) x 0.408 (RCB) = **495** ( $2\sigma \approx 44$ ) symptomatic seropositive individuals

→ Scaled to the same population as the control → **455** ( $2\sigma \approx 41$ )

**Control arm:**

163,838 participants x 0.0076 (SSP) x 0.399 (RCB) = **497** ( $2\sigma \approx 45$ ) symptomatic seropositive individuals

These formulas are correct if my contextual interpretation of the following (ambiguous) passage is correct: “Omitting symptomatic participants who did not consent to blood collection, symptomatic seroprevalence was 0.76% in control villages and 0.68% in the intervention villages. Because these numbers omit non-consenters, it is likely that the true rates of symptomatic seroprevalence are substantially higher (perhaps by 2.5 times, if non-consenters have similar seroprevalence to consenters).”

The difference,  $497 - 495 = 2$  individuals, is the number giving rise to Abaluck *et al.*'s difference in absolute symptomatic seroprevalence (SSP) of 0.0008. As such, given the expected sources of bias and measurement errors described herein, and given the size of this difference of only two (2) events, the SSP difference on increased masking in the treatment arm, reported by Abaluck *et al.*, cannot be taken as anything but unreliable.

The difference of “2 individuals” is 10 times smaller than the approximate ideal-statistical standard deviations ( $1\sigma$ ) of the numbers prior to taking their difference, for comparable size starting populations. This should give anyone pause.

If I pursue the calculation to obtain a prevalence ratio (PR), including 95 % confidence intervals,  $PR = 455 [414, 496] \div 497 [452, 542] = 0.92 [0.80, 1.04]$ , which is not statistically different from 1, and which gives a false impression of being borderline significant, from the purely ideal-statistical perspective.

Abaluck *et al.* report their results as: “Adjusting for baseline covariates, the intervention reduced symptomatic seroprevalence by 9.3% (adjusted prevalence ratio (aPR) = 0.91 [0.82, 1.00]; control prevalence 0.76%; treatment prevalence 0.68%).”

In fact, their bold assertion of a relative reduction in SSP of “9.3%”, without stating its ideal-statistical error, while ignoring all other-than-ideal-statistical errors, is a fiction.

It is also misleading for Abaluck *et al.* to present their percent relative reduction in SSP with two significant numbers (as “9.3%”): without “adjustment”, I calculate a percent relative reduction in SSP  $((497 - 455)/497)$  of **8.4 %  $\pm$  12.2 % ( $2\sigma$ )**, which is consistent with zero.

### *Oldest age group, 60+ years old, surgical masks only*

In their most surprising result, Abaluck *et al.* (2021) report a statistically significant three-significant-digit “34.7 %” relative decrease in symptomatic seroprevalence (from 1.03 % to 0.69 %, from control to treatment) among the 60+ years old age cohort, for surgical masks only in the treatment arm (their Figure 3).

Among other reasons, this result is surprising because all the many (>10) policy-grade randomized controlled trials (RCT) with lab-verified outcomes, for COVID-19 and other viral respiratory diseases, have found no statistically significant benefit from either surgical or N95 masks, in terms of transmission and infection. I have reviewed this context here: (Rancourt, 2021) (Rancourt, 2020a) (Rancourt, 2020b) (Rancourt, 2020c).

It is difficult to evaluate the said most surprising result of Abaluck *et al.* because the authors do not provide:

- the numbers of 60+ year olds in each group (control vs treatment with surgical masks)

- the fraction of distributed surgical masks to all distributed masks, in treatment-arm 60+ year olds
- the numbers of symptomatic 60+ year olds in each group (control vs treatment with surgical masks)
- the rate of consent to the blood test (RCB) in each group (control vs treatment with surgical masks)

On 13 September 2021, I emailed Dr. Abaluck directly and asked for these and other numbers of individuals: "... Basically, I am asking to know these 30 most basic numbers, only a few of which are already provided in your article. Can you or one of your co-authors provide these?" Dr. Abaluck responded the same day, as: "We will be posting replication instructions publicly in a few weeks and you'll be able to see all the data. If you can't find it in 3 weeks or so, please feel free to reach out again."

I note that Abaluck *et al.* (2021) do not provide ideal-statistical error estimates (confidence intervals) for any of their symptomatic seroprevalence numbers, for any group or arm. This leaves me with an impression of avoiding reporting estimated statistical uncertainties; while dealing solely with group to group differences and group to group relative changes of seroprevalence values having unreported error estimations.

Without the numbers for the 60+ year olds, it is impossible to definitively verify ideal-statistical uncertainty in the said most surprising result. Nonetheless, the needed uncertainties can be estimated using what is provided, by making reasonable assumptions for the missing information, as follows.

For this purpose: I assume the same RCB for 60+ year olds (control, surgical masks) as for all adults in the same arm. I assume that 16 % of adults in all groups are 60+ year olds (The World Factbook, for Bangladesh, 2020). I assume that 66.7 % of 60+ year olds receiving masks received surgical masks, equal to the cross-randomization fraction on a village basis (200/300).

I then estimate the numbers of symptomatic 60+ year olds having positive blood test results, as follows:

**Treatment group, 60+ year olds, surgical masks:**

178,288 participants x 0.16 (fraction 60+) x 0.667 (fraction surgical masks) x 0.0069 (SSP) x 0.408 (RCB)

= **54** ( $2\sigma \approx 15$ ) symptomatic seropositive 60+ year olds, surgical masks

→ Scaled to the same population as the control → **74** ( $2\sigma \approx 21$ )

**Control group, 60+ year olds:**

163,838 participants  $\times$  0.16 (fraction 60+)  $\times$  0.0103 (SSP)  $\times$  0.399 (RCB)  
 = **108** ( $2\sigma \approx 21$ ) symptomatic seropositive 60+ year olds, control

Thus I estimate that the two comparable numbers of symptomatic seropositive 60+ year old individuals overlap within their 95 % confidence intervals (74 [53, 95] (treatment); 108 [87, 129] (control)), from purely ideal-statistical considerations.

As a check, my numbers give a prevalence ratio (PR), 60+ year olds, surgical masks:

PR = 74 [53, 95] (treatment)  $\div$  108 [87, 129] (control) = 0.69 [0.45, 0.92],

which is close to the “adjusted” PR reported by Abaluck *et al.*:

aPR = 0.65 [0.46, 0.85].

Whereas this PR (aPR) for 60+ year olds and surgical masks has an appearance of being mathematically valid, it is not reliable, for the following reasons:

- i. The confidence interval is from purely ideal-statistical considerations. It is from the counting uncertainties alone, under ideal applicability assumptions. The main mathematical assumption is that each event or detection (of symptomatic seropositivity) is independent and random.
- ii. The actual (here estimated) absolute numbers of events or detections are small (54 and 108) and are therefore all the more susceptible to large errors from all sources, not just purely ideal-statistical counting errors. The smaller the cohorts, the greater the chance of contamination by unknown “baseline” factors, and the harder it is to secure a “balanced” comparison.
- iii. Observational bias error in reporting symptoms is expected, as explained above (impression of higher safety, unblind observers).
- iv. There is a built-in bias for resilience against infection in the treatment group, as explained above, which is expected to be strong, and is predicted to be strongest in the most elderly (stress-immune relation).
- v. There is an insufficiently large blood-testing rate of consent (RCB,  $\sim 40\%$ ), such that the non-randomized consent itself is therefore susceptible to bias.
- vi. The laboratory test is not specific to SARS-CoV-2, is not validated for Bangladesh, and is susceptible to large occurrences of “undetermined” or “equivocal” readings, as explained above, all of which make it susceptible to bias in whatever it is detecting or not detecting.

- vii. Many factors may be highly imbalanced between the treatment and control arms, which are not known or controlled in the study. These factors include infections, conditions or pathologies that have possible or likely cross-reactivity in the serology test, as explained above. This potential is probably higher in the most elderly, who are often afflicted with several co-conditions.
- viii. There is a large (50 %) imbalance in “baseline symptomatic seroprevalence rate”: 0.00002 (treatment), 0.00003 (control) (their “Table 1: Balance Tests (Individual-Level)” and “Table A3: Balance Tests (Village-Level)”). Abaluck *et al.* do not explain “rate” or discuss or attempt to interpret this apparently fundamental difference. This imbalance may indicate different immune histories or different immune health of the individuals or different pathogenic environments in the control and treatment arms.
- ix. There may be unaccounted or unknown correlations or clustering that vitiate the assumption of ideal-statistical independence and randomness. For example, a 60+ year old may have a higher-than-otherwise (higher than random) probability of being symptomatic seropositive if another 60+ year old in the same household is or recently was symptomatic seropositive, and so on. After all, the study includes all adults per participating household, rather than the common/standard study design of having independent participants. (This means that the method of calculation of confidence intervals for this study design, looking at individuals, is itself strictly invalid; as are all individual-base prevalence and prevalence-ratio results.)
- x. There may be hidden co-factors that produce COVID-like symptoms and give cross-reactivity in the serology test. The door is wide open for this possibility since the COVID-19 symptoms are rather generic and the serology test is far from having been evaluated to be specific for SARS-CoV-2, as show above. The small absolute numbers of events or detections (54 and 108) allow such co-factors (one or several) to be accidentally different to a large extent in the two groups.
- xi. Symptomatic seropositivity for COVID-19 was not confirmed by clinical diagnosis; and symptomatic seroprevalence (SSP) was not validated by hospitalization data or mortality or prescription data or absenteeism, etc. Abaluck *et al.* give no information about number and severity of symptoms, but instead use a binary threshold of “symptomatic”. What was comparative symptomatology (severity, etc.) in the small numbers for the two groups (54 and 108)?

## Conclusion

The Abaluck *et al.* (2021) study is an extreme case in which a Bayesian analysis of the impact of foreseeable potential bias and measurement uncertainty would confirm that their results are false, but the sophisticated demonstration is hardly necessary (Ioannidis, 2005) (Greenland, 2006).

In technical language, it is a case of “garbage in, garbage out”, not to mention the fundamental design flaws including using households while extracting individual-base results, and applying impactful super-treatment interventions to the treatment arm.

If this is the new “gold-standard clinical trial” ([according to Nature](#)) then the value of gold has plummeted to that of lead.

And see: Appendix A.

## References

Abaluck *et al.* (2021) “The Impact of Community Masking on COVID-19: A Cluster-Randomized Trial in Bangladesh”. *Innovations for Poverty Action (IPA)*. 01 September 2021. Working Paper, dated 31 August 2021. <https://www.poverty-action.org/publication/impact-community-masking-covid-19-cluster-randomized-trial-bangladesh> -- or -- <https://elischolar.library.yale.edu/egcenter-discussion-paper-series/1086/>

Ader and Cohen. (1993) “Psychoneuroimmunology: Conditioning and Stress”. *Annual Review of Psychology* 1993 44:1, 53-85. <https://pubmed.ncbi.nlm.nih.gov/8434895/>

Cohen *et al.* (2007) “Psychological Stress and Disease”. *JAMA*, 298(14), pp. 1685–1687. doi: 10.1001/jama.298.14.1685. <https://pubmed.ncbi.nlm.nih.gov/17925521/>

Cohen *et al.* (1997b) “Social Ties and Susceptibility to the Common Cold”. *JAMA*, 277(24), pp. 1940–1944. doi: 10.1001/jama.1997.03540480040036. <https://pubmed.ncbi.nlm.nih.gov/9200634/>

Cohen *et al.* (1997a) “Chronic Social Stress, Social Status, and Susceptibility to Upper Respiratory Infections in Nonhuman Primates”. *Psychosomatic Medicine*: May/June 1997 - Volume 59 - Issue 3 - p 213-221.

[https://kilthub.cmu.edu/articles/journal\\_contribution/Chronic\\_Social\\_Stress\\_Social\\_Status\\_and\\_Susceptibility\\_to\\_Upper\\_Respiratory\\_Infections\\_in\\_Nonhuman\\_Primates/6613937/files/12106595.pdf](https://kilthub.cmu.edu/articles/journal_contribution/Chronic_Social_Stress_Social_Status_and_Susceptibility_to_Upper_Respiratory_Infections_in_Nonhuman_Primates/6613937/files/12106595.pdf)

Cohen *et al.* (1991) “Psychological Stress and Susceptibility to the Common Cold”. *New England Journal of Medicine*. Massachusetts Medical Society, 325(9), pp. 606–612. doi: 10.1056/NEJM199108293250903. <https://pubmed.ncbi.nlm.nih.gov/1713648/>

Dhabhar. (2014) “Effects of stress on immune function: the good, the bad, and the beautiful”. *Immunologic Research*. 2014 May; 58(2-3): 193-210. doi: 10.1007/s12026-014-8517-0. PMID: 24798553. (cited >800) <https://link.springer.com/article/10.1007%2Fs12026-014-8517-0>

FDA (Test Performance, 2021) “EUA Authorized Serology Test Performance”. Content current as of: 08/18/2021; accessed on 14 September 2021. <https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance>

FNLCR (2021) “Serology Test Evaluation Report for “SCoV-2 Detect™ IgG ELISA” from InBios International Inc.”. *Frederick National Laboratory for Cancer Research*. 13 July 2021. [https://www.accessdata.fda.gov/cdrh\\_docs/presentations/maf/maf3315-a001.pdf](https://www.accessdata.fda.gov/cdrh_docs/presentations/maf/maf3315-a001.pdf)

Greenland. (2006) “Bayesian perspectives for epidemiological research: I. Foundations and basic methods, *International Journal of Epidemiology*, Volume 35, Issue 3, June 2006, Pages 765–775, <https://doi.org/10.1093/ije/dyi312>

InBios (IFU LBL-0113-03, 2021) “InBios - SCoV-2 Detect™ IgG ELISA - Instructions for Use”. *InBios/FDA*. COVE-G EUA/CE SCoV-2 Detect™ IgG ELISA. Insert Part No. 900255-03. Effective Date: 05/19/2021. <https://inbios.com/wp-content/uploads/2021/05/LBL-0113-03-EUA-CE-SCoV-2-Detect-IgG-ELISA-product-insert-English.pdf>

Ioannidis. (2005) “Why Most Published Research Findings Are False”. *PLoS Med* 2(8): e124. <https://doi.org/10.1371/journal.pmed.0020124>

Islam *et al.* (2013) “An epidemiological overview of malaria in Bangladesh”. *Travel Med Infect Dis*. 2013 Jan-Feb;11(1):29-36. doi: 10.1016/j.tmaid.2013.01.004. Epub 2013 Feb 21. PMID: 23434288. [https://scholar.harvard.edu/files/naz/files/epidemiology\\_malaria\\_bangladesh\\_travel\\_med\\_inf\\_dis\\_2013.pdf](https://scholar.harvard.edu/files/naz/files/epidemiology_malaria_bangladesh_travel_med_inf_dis_2013.pdf)



Kaufman *et al.* (2021) "Insights from Patterns of SARS-CoV-2 Immunoglobulin G Serology Test Results from a National Clinical Laboratory, United States, March–July 2020". *Population Health Management*, 24(S1), S35–S42. <https://doi.org/10.1089/pop.2020.0256> ----  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7875137/>

Peters *et al.* (2021) "To stress or not to stress: Brain-behavior-immune interaction may weaken or promote the immune response to SARS-CoV-2". *Neurobiology of Stress*, Volume 14, 100296. ISSN 2352-2895. <https://doi.org/10.1016/j.ynstr.2021.100296>.

Prenderville *et al.* (2015) "Adding fuel to the fire: the impact of stress on the ageing brain". *Trends in Neurosciences*, 38(1), pp. 13–25. doi: 10.1016/j.tins.2014.11.001.  
<https://pubmed.ncbi.nlm.nih.gov/25705750/>

Rancourt. (2021) "Review of scientific reports of harms caused by face masks, up to February 2021". *ResearchGate*. 22 February 2021. DOI: 10.13140/RG.2.2.14294.37448.  
<https://archive.vn/OL5ji>

Rancourt. (2020a) "Measures do not prevent deaths, transmission is not by contact, masks provide no benefit, vaccines are inherently dangerous: Review update of recent science relevant to COVID-19 policy". *ResearchGate*. 28 December 2020.  
DOI: 10.13140/RG.2.2.21706.18885. <https://archive.ph/F5xqy>

Rancourt. (2020b) "Face masks, lies, damn lies, and public health officials: "A growing body of evidence"". *ResearchGate*. 03 August 2020. DOI: 10.13140/RG.2.2.25042.58569.  
<https://archive.ph/BiUhb>

Rancourt. (2020c) "Masks Don't Work: A review of science relevant to COVID-19 social policy". *ResearchGate*. 11 April 2020. DOI: 10.13140/RG.2.2.14320.40967/1. <https://archive.ph/RuA5z>  
([article history](#))

Sapolsky. (2005) "The Influence of Social Hierarchy on Primate Health", *Science*, 29 April 2005, vol. 308, pages 648-652. DOI: 10.1126/science.1106477.  
<https://www.pinniped.net/sapolsky2005.pdf>

Usher-Smith *et al.* (2016) "The spectrum effect in tests for risk prediction, screening, and diagnosis". *BMJ* 2016; 353 :i3139 doi:10.1136/bmj.i3139.  
<https://www.bmj.com/content/353/bmj.i3139>

Yadouleton *et al.* (2021) "Limited Specificity of Serologic Tests for SARS-CoV-2 Antibody Detection, Benin". *Emerg Infect Dis*. 2021;27(1):233-237.  
<https://doi.org/10.3201/eid2701.203281>

## Appendix A: Media reviews of the Abaluck *et al.* (2021) mask study

A few features made me suspicious of the Abaluck *et al.* (2021) study. The first was the high octane media campaign, followed by my noting the presence of clearly false statements in the media articles.

Another was the self-serving and incomplete description of the context of face mask efficacy studies, made by the authors themselves, in-effect ignoring all existing policy-grade trials that find no detectable advantage to mask wearing, in terms of transmission and infection. Abaluck *et al.* summarise as: “Inspired by the growing body of scientific evidence that face masks can slow the spread of the disease and save lives [refs], we conducted...”; and they never attempt to reconcile their surprising results with the existing science.

I infer that Abaluck *et al.* may self-justify in-effect ignoring all past work by distinguishing “source control” and “protective effect” of face masks? They sate: “First, unlike technologies with primarily private benefits, mask adoption is likely to yield especially large benefits at the community-level.” This concept of “the one-way mask” is not based of any empirical evidence of actual person-to-person transmission. It also seems contrary to mechanistic expectations. If masks filter relevant particles, then they should filter them in both directions, both inhaling and exhaling. Exhaling is towards the outside environment, whereas inhaling is directly towards the respiratory tract tissue that is the target of the pathogen. If face masks are “one-way” then it should be the other way.

Here is a sample of the media reports:

— *Nature* | News | 09 September 2021 | [“Face masks for COVID pass their largest test yet”](#)

Face masks protect against COVID-19. That’s the conclusion of a [gold-standard clinical trial](#) in Bangladesh, which backs up the findings of hundreds of previous observational and laboratory studies.[ref].

Critics of mask mandates have cited the lack of relevant randomized clinical trials, which assign participants at random to either a control group or an intervention group. But the latest finding is based on a randomized trial involving nearly 350,000 people across rural Bangladesh. The study’s authors found that surgical masks — but not cloth

masks — reduced transmission of SARS-CoV-2 in villages where the research team distributed face masks and promoted their use.

“This really should be the end of the debate,” says Ashley Styczynski, an infectious-disease researcher at Stanford University in California and a co-author of the preprint describing the trial. The research “takes things a step further in terms of scientific rigour”, says Deepak Bhatt, a medical researcher at Harvard Medical School in Boston, Massachusetts, who has published research on masking. ...

— *Stanford Medicine* | News Center | 01 September 2021 | [“Surgical masks reduce COVID-19 spread, large-scale study shows”](#)

The findings were released Sept. 1 on the Innovations for Poverty Action website, prior to their publication in a scientific journal, because the information is considered of pressing importance for public health as the pandemic worsens in many parts of the world.

“We now have evidence from a randomized, controlled trial that mask promotion increases the use of face coverings and prevents the spread of COVID-19,” said Stephen Luby, MD, professor of medicine at Stanford. “This is the gold standard for evaluating public health interventions. Importantly, this approach was designed to be scalable in lower- and middle-income countries struggling to get or distribute vaccines against the virus.”

— *The Washington Post* | 01 September 2021 | [“Massive randomized study is proof that surgical masks limit coronavirus spread, authors say”](#)

The authors of a study based on an enormous randomized research project in Bangladesh say their results offer the best evidence yet that widespread wearing of surgical masks can limit the spread of the coronavirus in communities.

The preprint paper, which tracked more than 340,000 adults across 600 villages in rural Bangladesh, is by far the largest randomized study on the effectiveness of masks at limiting the spread of coronavirus infections.

Its authors say this provides conclusive, real-world evidence for what laboratory work and other research already strongly suggest: mask-wearing can have a significant impact on limiting the spread of symptomatic covid-19, the disease caused by the virus.

“I think this should basically end any scientific debate about whether masks can be effective in combating covid at the population level,” Jason Abaluck, an economist at Yale who helped lead the study, said in an interview, calling it “a nail in the coffin” of the arguments against masks.

— *NBC News* | 01 September 2021 | [“Largest study of masks yet details their importance in fighting Covid-19”](#)

A study involving more than 340,000 people in Bangladesh offers some of the strongest real-world evidence yet that mask use can help communities slow the spread of Covid-19.

The research, conducted across 600 villages in rural Bangladesh, is the largest randomized trial to demonstrate the effectiveness of surgical masks, in particular, to curb transmission of the coronavirus. Though previous, smaller studies in laboratories and hospitals have shown that masks can help prevent the spread of Covid, the new findings demonstrate that efficacy in the real world — and on an enormous scale.

“This is really solid data that combines the control of a lab study with real-life actions of people in the world to see if we can get people to wear masks, and if the masks work,” said Laura Kwong, an assistant professor of environmental health sciences at the University of California, Berkeley, and one of the co-authors of the study.

— *Berkeley Public Health* | 01 September 2021 (undated) | [“Largest study of its kind finds face masks reduce COVID-19”](#)

Wearing face masks, particularly surgical masks, is truly effective in reducing the spread of COVID-19 in community settings, finds a new study led by researchers from Yale University, Stanford Medical School, the University of California, Berkeley, and the nonprofit Innovations for Poverty Action (IPA). ...

“These results suggest that we could prevent unnecessary death and disease if we get people to wear high-performance masks, such as surgical masks, in schools, workplaces, shopping centers, places of worship and other indoor spaces,” said study co-author Laura Kwong, an assistant professor of environmental health sciences at Berkeley’s School of Public Health.

— *The Atlantic* | 04 September 2021 | [“The Masks Were Working All Along”](#)

Now we have definitive proof that masks really are effective.

... Their conclusion? Masks work, period. Surgical masks are particularly effective at preventing coronavirus transmission. And community-wide mask wearing is excellent at protecting older people, who are at much higher risk of severe illness from COVID-19.

— *Yale Daily News* | 13 September 2021 | "[First randomized trial on masking affirms efficacy, Yale study says](#)"

... The 300,000-person study was the first randomized trial on mask efficacy.

Yale professors of economics Ahmed Mushfiq Mobarak and Jason Abaluck, alongside a team of researchers from Stanford University and the University of California at Berkeley, conducted a cluster-randomized trial in rural Bangladesh that tested the intervention of community-level masking promotion from November 2020 to April 2021. ...

"A lot of conversation around mask usage previously had been that there had never been a randomized, controlled trial that demonstrated that masks were effective in both interrupting and preventing disease," said Stephen Luby, professor of infectious diseases at Stanford University and a coauthor of the study. "This really was a gold standard trial and was able to demonstrate just that."

— *WebMD Health News* | 07 September 2021 | "[Large Study Confirms Masks Work to Limit COVID-19 Spread](#)"

The study demonstrates the power of careful investigation and offers a host of lessons about mask wearing that will be important worldwide. ...

"What we really were able to achieve is to demonstrate that masks are effective against COVID-19, even under a rigorous and systematic evaluation that was done in the throes of the pandemic," said Ashley Styczynski, MD, who was an infectious disease fellow at Stanford University when she collaborated on the study with other colleagues at Stanford, Yale, and Innovations for Poverty Action (IPA), a large research and policy nonprofit organization that currently works in 22 countries.

## My competence to review science about COVID-19

I am a former tenured Full Professor of Physics, University of Ottawa, Canada. Full Professor is the highest academic rank. During my 23-year career as a university professor, I developed new courses and taught over 2000 university students, at all levels, and in three different faculties (Science, Engineering, Arts). I supervised more than 80 junior research terms or degrees at all levels from post-doctoral fellow to graduate students to NSERC undergraduate researchers. I headed an internationally recognized interdisciplinary research laboratory, and attracted significant research funding for two decades.

I have been an invited plenary, keynote, or special session speaker at major scientific conferences some 40 times. I have published over 100 research papers in leading peer-reviewed scientific journals, in the areas of physics, chemistry, geology, bio-geochemistry, measurement science, soil science, and environmental science.

My scientific h-index impact factor is 41, and my articles have been cited more than 5,000 times in peer-reviewed scientific journals (profile at Google Scholar: <https://scholar.google.ca/citations?user=1ChsRsQAAAAJ> ).

My personal knowledge and ability to evaluate the facts in this article are grounded in my education, research, training and experience, as follows ( <https://denisrancourt.ca/> ):

- i. *Regarding environmental nanoparticles.* Viral respiratory diseases are transmitted by the smallest size-fraction of virion-laden aerosol particles, which are reactive environmental nanoparticles. Therefore, the chemical and physical stabilities and transport properties of these aerosol particles are the foundation of the dominant contagion mechanism through air. My extensive work on reactive environmental nanoparticles is internationally recognized, and includes: precipitation and growth, surface reactivity, agglomeration, surface charging, phase transformation, settling and sedimentation, and reactive dissolution. In addition, I have taught the relevant fluid dynamics (air is a compressible fluid), and gravitational settling at the university level, and I have done industrial-application research on the technology of filtration (face masks are filters).
- ii. *Regarding molecular science, molecular dynamics, and surface complexation.* I am an expert in molecular structures, reactions, and dynamics, including molecular complexation to biotic and abiotic surfaces. These processes are the basis of viral attachment, antigen attachment, molecular replication, attachment to mask fibers, particle charging, loss and growth in aerosol particles, and all such phenomena involved in viral transmission and infection, and in protection measures. I taught quantum mechanics at the advanced university level for many years, which is the fundamental theory of atoms, molecules and substances; and in my published research I developed X-ray diffraction theory and methodology for characterizing small material particles.
- iii. *Regarding statistical analysis methods.* Statistical analysis of scientific studies, including robust error propagation analysis and robust estimates of bias, sets the limit of what reliably can be inferred from any observational study, including randomized controlled trials in medicine, and including field measurements during epidemics. I am an expert in error analysis and statistical analysis of complex data, at the research level in many areas of science. Statistical analysis methods are the basis of medical research.
- iv. *Regarding mathematical modelling.* Much of epidemiology is based on mathematical models of disease transmission and evolution in the population. I have research-level knowledge and experience with predictive and exploratory mathematical models and simulation methods. I have expert knowledge related to parameter uncertainties and parameter dependencies in such models.

I have made extensive simulations of epidemiological dynamics, using standard compartmental models (SIR, MSIR) and new models.

- v. *Regarding measurement methods.* In science there are five main categories of measurement methods: (1) spectroscopy (including nuclear, electronic and vibrational spectroscopies), (2) imaging (including optical and electron microscopies, and resonance imaging), (3) diffraction (including X-ray and neutron diffractions, used to elaborate molecular, defect and magnetic structures), (4) transport measurements (including reaction rates, energy transfers, and conductivities), and (5) physical property measurements (including specific density, thermal capacities, stress response, material fatigue...). I have taught these measurement methods in an interdisciplinary graduate course that I developed and gave to graduate (M.Sc. and Ph.D.) students of physics, biology, chemistry, geology, and engineering for many years. I have made fundamental discoveries and advances in areas of spectroscopy, diffraction, magnetometry, and microscopy, which have been published in leading scientific journals and presented at international conferences. I know measurement science, the basis of all sciences, at the highest level.