

EYE ON THE NEWS

Do Masks Work?

A review of the evidence

Jeffrey H. Anderson

August 11, 2021

“Seriously people—STOP BUYING MASKS!” So tweeted then–surgeon general Jerome Adams on February 29, 2020, adding, “They are NOT effective in preventing general public from catching #Coronavirus.” Two days later, Adams said, “Folks who don’t know how to wear them properly tend to touch their faces a lot and actually can increase the spread of coronavirus.” Less than a week earlier, on February 25, public-health authorities in the United Kingdom had published guidance that masks were unnecessary even for those providing community or residential care: “During normal day-to-day activities facemasks do not provide protection from respiratory viruses, such as COVID-19 and do not need to be worn by staff.” About a month later, on March 30, World Health Organization (WHO) Health Emergencies Program executive director Mike Ryan said that “there is no specific evidence to suggest that the wearing of masks by the mass population has any particular benefit.” He added, “In fact there’s some evidence to suggest the opposite” because of the possibility of not “wearing a mask properly or fitting it properly” and of “taking it off and all the other risks that are otherwise associated with that.”

Surgical masks were designed to keep medical personnel from inadvertently infecting patients’ wounds, not to prevent the spread of viruses. Public-health officials’ advice in the early days of Covid-19 was consistent with that understanding. Then, on April 3, 2020, Adams announced that the CDC was changing its guidance and that the general public should hereafter wear masks whenever sufficient social distancing could not be maintained.

Fast-forward 15 months. Rand Paul has been suspended from YouTube for a week for saying, “Most of the masks you get over the counter don’t work.” Many cities across the country, following new CDC guidance handed down amid a spike in cases

nationally caused by the Delta variant, are once again mandating indoor mask-wearing for everyone, regardless of inoculation status. The CDC further recommends that all schoolchildren and teachers, even those who have had Covid-19 or have been vaccinated, should wear masks.

The CDC asserts this even though its own statistics show that Covid-19 is not much of a threat to schoolchildren. Its numbers show that more people under the age of 18 died of influenza during the [2018–19 flu season](#)—a season of “moderate severity” that lasted eight months—than have [died of Covid-19](#) across more than 18 months. What’s more, the CDC [says](#) that out of every 1,738 Covid-19-related deaths in the U.S. in 2020 and 2021, just one has involved someone under 18 years of age; and out of every 150 deaths of someone under 18 years of age, just one has been Covid-related. Yet the CDC declares that schoolchildren, who learn in part from communication conveyed through facial expressions, should nevertheless hide their faces—and so should their teachers.

How did mask guidance change so profoundly? Did the medical research on the effectiveness of masks change—and in a remarkably short period of time—or just the guidance on wearing them?

Since we are constantly told that the CDC and other public-health entities are basing their recommendations on science, it’s crucial to know what, specifically, has been found in various medical studies. Significant choices about how our republic should function cannot be made on the basis of science alone—they require judgment and the weighing of countless considerations—but they must be informed by knowledge of it.

In truth, the CDC’s, U.K.’s, and WHO’s earlier guidance was much more consistent with the best medical research on masks’ effectiveness in preventing the spread of viruses. That research suggests that Americans’ many months of mask-wearing has likely provided little to no health benefit and might even have been counterproductive in preventing the spread of the novel coronavirus.

It’s striking how much the CDC, in marshalling evidence to justify its revised mask guidance, studiously avoids mentioning randomized controlled trials. RCTs are uniformly regarded as the gold standard in medical research, yet the CDC basically ignores them apart from disparaging certain ones that particularly contradict the agency’s position. In a [“Science Brief”](#) highlighting studies that “demonstrate that mask wearing reduces new infections” and serving as the main public justification for its mask guidance, the CDC provides a helpful matrix of 15 studies—none RCTs.

The CDC instead focuses strictly on observational studies completed after Covid-19 began. In general, observational studies are not only of lower quality than RCTs but also are more likely to be politicized, as they can inject the researcher's judgment more prominently into the inquiry and lend themselves, far more than RCTs, to finding what one wants to find.

A particular favorite of the CDC's, so much so that the agency put out a glowing press release on it and continues to give it pride of placement in its brief, is an observational (specifically, cohort) [study](#) focused on two Covid-positive hairstylists at a beauty salon in Missouri. The two stylists, who were masked, provided services for 139 people, who were mostly masked, for several days after developing Covid-19 symptoms. The 67 customers who subsequently chose to get tested for the coronavirus tested negative, and none of the 72 others reported symptoms.

This study has major limitations. For starters, any number of the 72 untested customers could have had Covid-19 but been asymptomatic, or else had symptoms that they chose not to report to the Greene County Health Department, the entity doing the asking. The apparent lack of spread of Covid-19 could have been a result of good ventilation, good hand hygiene, minimal coughing by the stylists, or the fact that stylists generally, as the researchers note, "cut hair while clients are facing away from them." The researchers also observe that "viral shedding" of the coronavirus "is at its highest during the 2 to 3 days before symptom onset." Yet no customers who saw the stylists when they were at their most contagious were tested for Covid-19 or asked about symptoms. Most importantly, this study does not have a control group. Nobody has any idea how many people, if any, would have been infected had no masks been worn in the salon. Late last year, at a gym in Virginia in which people apparently did not wear masks most of the time, a trainer tested positive for the coronavirus. As CNN [reported](#), the gym contacted everyone whom the trainer had coached before getting sick—50 members in all—"but not one member developed symptoms." Clearly, this doesn't prove that *not* wearing masks prevents transmission.

Another CDC-highlighted [study](#), by Rader et al., invited people across the country to answer a survey. The low (11 percent) response rate—including about twice as many women as men—indicated that the mix of respondents was hardly random. The study found that "a high percentage of self-reported face mask-wearing is associated with a higher probability of transmission control," and "the highest percentage of reported mask wearers" are found, unsurprisingly, "along the coasts and southern border, and in large urban areas." However, as the researchers note, "It is difficult to disentangle individuals' engagement in mask-wearing from their adoption of other preventive hygiene practices, and mask-wearing might serve as a proxy for other risk

avoidance behaviors not queried.” Moreover, achieving greater “transmission control” is not remotely the same thing as ensuring fewer deaths. For example, per capita, Utah is in the top ten in the nation in Covid-19 cases and the bottom ten in Covid-19 deaths, while Massachusetts is in the bottom half in cases and the top five in deaths.

An additional observational study, but one that the CDC does not reference in its brief, is a large, international Bayesian study by Leech, et al. It finds that mask-wearing by 100 percent of the population “corresponds to” a 24.6 percent reduction in transmission of the novel coronavirus. Mask mandates correspond to no decrease in transmission: “For mandates we see no reduction: 0.0 percent.” Like all observational studies, however, this study is ill-equipped to show causation, to separate out the effects of just one variable from among other, frequently related, ones.

Mask supporters often claim that we have no choice but to rely on observational studies instead of RCTs, because RCTs cannot tell us whether masks work or not. But what they really mean is that they don’t like what the RCTs show.

The randomized controlled trial dates, in a sense, to 1747, when Royal Navy surgeon James Lind divided seamen suffering from similar cases of scurvy into six pairs and tried different methods of treatment on each. Lind writes, “The consequence was, that the most sudden and visible good effects were perceived from the use of oranges and lemons.”

The RCT eventually became firmly established as the most reliable way to test medical interventions. The following passage, from Abdelhamid Attia, an M.D. and professor of obstetrics and gynecology at Cairo University in Egypt, conveys its dominance:

The importance of RCTs for clinical practice can be illustrated by its impact on the shift of practice in hormone replacement therapy (HRT). For decades HRT was considered the standard care for all postmenopausal, symptomatic and asymptomatic women. Evidence for the effectiveness of HRT relied always on observational studies[,] mostly cohort studies. But a single RCT that was published in 2002 . . . has changed clinical practice all over the world from the liberal use of HRT to the conservative use in selected symptomatic cases and for the shortest period of time. In other words, one well conducted RCT has changed the practice that relied on tens, and probably hundreds, of observational studies for decades.

A randomized controlled trial divides participants into different groups on a randomized basis. At least one group receives an “intervention,” or treatment, that is generally tested against a control group not receiving the intervention. The twofold strength of an RCT is that it allows researchers to isolate one variable—to test whether a given intervention causes an intended effect—while at the same time making it very hard for researchers to produce their own preferred outcomes.

This is true at least so long as an RCT’s findings are based on “intention-to-treat” analysis, whereby all participants are kept in the treatment group to which they were originally assigned and none are excluded from the analysis, regardless of whether they actually received the intended treatment. Eric McCoy, an M.D. at the University of California, Irvine, [explains](#) that intention-to-treat analysis avoids bias and “preserves the benefits of randomization, which cannot be assumed when using other methods of analysis.”

Such other methods of analysis include subgroup, multivariable, and per-protocol analysis. Subgroup analysis is susceptible to “cherry-picking”—as researchers hunt for anything showing statistical significance—or to being swayed by random chance. In one famous example, aspirin was found to help prevent fatal heart attacks, but not in the subgroups where patients’ astrological signs were Gemini or Libra.

“Multivariable analysis,” [writes](#) Marlies Wakkee, an M.D. and Ph.D. at Erasmus University Medical Center in the Netherlands, “only adjusts for measured confounding”—that which a researcher decides is worth examining. (Confounders are extra variables that affect the analysis; for example, eating ice cream may be found to correlate with sunburns, but heat is a confounding variable influencing both.) She adds, “This is a significant difference compared to randomized controlled trials, where the randomization process results in an equal distribution of all potential confounders, known and unknown.”

Per-protocol analysis departs from randomization by basically allowing participants to self-select into, or out of, an intervention group. McCoy writes, “Empirical evidence suggests that participants who adhere [to research protocols] tend to do better than those who do not adhere, regardless of assignment to active treatment or placebo.” In other words, per-protocol analysis is more likely to suggest that an intervention, even a fake one, worked. Of these three departures from intention-to-treat analysis, per-protocol analysis is perhaps the most extreme.

With these different methods of analysis in mind, it becomes easier to evaluate the 14 RCTs, conducted around the world, that have tested the effectiveness of masks in

reducing the transmission of respiratory viruses. Of these 14, the two that have directly tested “source control”—the oft-repeated claim that wearing a mask benefits others—are a good place to start.

A 2016 [study](#) in Beijing by MacIntyre, et al. that claimed to find a possible benefit of masks did not prove very informative, as only one person in the control group—and one in the mask group—developed a laboratory-confirmed infection. Much more illuminating was a 2010 [study](#) in France by Canini, et al., which randomly placed sick people, or “index patients,” and their household contacts together into either a mask group or a no-mask control group. The authors “observed a good adherence to the intervention,” meaning that the index patients generally wore the furnished three-ply masks as intended. (No one else was asked to wear them.) Within a week, 15.8 percent of household contacts in the no-mask control group and 16.2 percent in the mask group developed an “influenza-like illness” (ILI). So, the two groups were essentially dead even, with the sliver of an advantage observed in the control group not being statistically significant. The authors write that the study “should be interpreted with caution since the lack of statistical power prevents us to draw formal conclusion regarding effectiveness of facemasks in the context of a seasonal epidemic.” However, they state unequivocally, “In various sensitivity analyses, we did not identify any trend in the results suggesting effectiveness of facemasks.”

With the two RCTs that directly tested source control providing essentially no support for the claim that wearing a mask benefits others, what about RCTs that test the combination of source control and wearer protection? By dividing participants into a hand-hygiene group, a hand-hygiene group that also wore masks, and a control group, three RCTs allow us to see whether the addition of masks (worn both by the sick person and others) provided any benefit over hand hygiene alone.

A 2010 [study](#) by Larson, et al. in New York found that those in the hand-hygiene group were less likely to develop any symptoms of an upper respiratory infection (42 percent experienced symptoms) than those in the mask-plus-hand-hygiene group (61 percent). This statistically significant finding suggests that wearing a mask actually undermines the benefits of hand hygiene.

A multivariable analysis of this same study found a significant difference in secondary attack rates (the rate of transmission to others) between the mask-plus-hands group and the control group. On this basis, the authors maintain that mask-wearing “should be encouraged during outbreak situations.” However, this multivariable analysis also found significantly lower rates in crowded homes—“i.e., more crowded households had less transmission”—which tested at a higher

confidence level. Thus, to the extent that this multivariable analysis provided any support for masks, it provided at least as much support for crowding.

Two other studies found no statistically significant differences between their mask-plus-hands and hands-only groups. A 2011 [study](#) in Bangkok by Simmerman, et al. observed very similar results for both groups. A CDC-funded 2009 [study](#) in Hong Kong by Cowling, et al. observed that the hands-only group generally did better than the mask-plus-hands group, but not to a statistically significant degree. Subgroup analysis by Cowling, et al., limited to interventions started within 36 hours of the onset of symptoms, found that the mask-plus-hands group beat the control group to a statistically significant degree in one measure, while the hands-only group beat the control group to a statistically significant degree in two measures. Summarizing this study, Canini writes that “no additional benefit was observed when facemask [use] was added to hand hygiene by comparison with hand hygiene alone.”

So, if masks don't improve on hand hygiene alone, what about masks versus nothing?

Various RCTs have studied this question, with evidence of masks' effectiveness proving sparse at best. Aside from a 2009 [study](#) in Japan by Jacobs, et al.—which found that those in the mask group were significantly more likely to experience headaches and that “face mask use in health care workers has not been demonstrated to provide benefit”—only two RCTs have produced statistically significant findings in intention-to-treat analysis, and one of those studies contradicted itself.

The previously mentioned 2011 study in Bangkok by Simmerman, et al. found that the secondary attack rate of ILI was twice as high in the mask-plus-hand-hygiene group (18 percent) as in the control group (9 percent), a statistically significant difference. (The ILI rate was 17 percent in the hand-hygiene-only group.) Finding essentially the same thing in multivariable analysis, the researchers wrote that, relative to the control group, the odds ratios for both the mask-plus-hands group and the hands-only group “were twofold in the opposite direction from the hypothesized protective effect.”

Subsequently, a small 2014 [study](#)—with 164 participants—by Barasheed, et al. of Australian pilgrims in Saudi Arabia, staying in close quarters in tents, found that significantly fewer people in the mask group developed an ILI than in the control group (31 percent to 53 percent). Unlike the exact fever specifications utilized in other RCTs, however, this study accepted self-reporting of “subjective” fever in determining whether someone had an ILI. Lab tests revealed opposite results, with

twice as many participants having developed respiratory viruses in the mask group as in the control group. These lab-test findings were not statistically significant; still, the lab tests' greater reliability makes it far from clear that the masks in this study provided any genuine benefit.

Other RCTs found no statistically significant benefit from masks in intention-to-treat analysis. A 2008 pilot [study](#) by Cowling et al. in Hong Kong observed that secondary attack rates, using the CDC's definition of ILI, were twice as high in the mask group (8 percent) as in the hand hygiene (4 percent) or control (4 percent) groups, but these observed differences were not statistically significant.

Other methods of analysis, deviating from intention-to-treat analysis, found the following.

A per-protocol [analysis](#) of a 2009 study in Sydney by MacIntyre, et al. found a significant effect when combining the surgical-mask group with a group wearing N95 hospital respirators. However, the authors write, a "causal link cannot be demonstrated because adherence was not randomized."

In subgroup analysis of [2010](#) and [2012](#) studies in Michigan by Aiello, et al., limited to the final several weeks of the respective studies, each study's mask-plus-hands group had significantly lower rates of ILI than its control group, while its mask-only group did not. In 2010, the results for the mask-only group also hinted at a slight benefit, reducing ILI by an observed (but not statistically significant) 8 percent to 10 percent. In 2012, the authors concluded, "Masks alone did not provide a benefit." They nevertheless recommended the combination of mask use and hand hygiene, despite not having tested whether that combination works better than hand hygiene alone.

A multivariable analysis of a smallish (218 participants) 2012 [study](#) in Germany by Suess, et al. found that combining the mask group and mask-plus-hands group, while limiting analysis to interventions begun within 48 hours, produced a finding of significantly lower levels of lab-confirmed influenza (but not of ILI) in that combined group (but not in either group separately). The authors, from Berlin, recommended masking and hand hygiene, while opining, "Concerns about acceptability and tolerability of the interventions should not be a reason against their recommendation."

The only RCT to test mask-wearing's specific effectiveness against Covid-19 was a 2020 [study](#) by Bundgaard, et al. in Denmark. This large (4,862 participants) RCT divided people between a mask-wearing group (providing "high-quality" three-layer

surgical masks) and a control group. It took place at a time (spring 2020) when Denmark was encouraging social distancing but not mask use, and 93 percent of those in the mask group wore the masks at least “predominately as recommended.” The study found that 1.8 percent of those in the mask group and 2.1 percent of those in the control group became infected with Covid-19 within a month, with this 0.3-point difference not being statistically significant.

This study—the first RCT on Covid-19 transmission—apparently had difficulty getting published. After the study’s eventual publication, Vinay Prasad, an M.D. at the University of California, San Francisco, described it as “thoughtful,” “useful,” and “well done,” but noted (with criticism), “Some have turned to social media to ask why a trial that may diminish enthusiasm for masks and may be misinterpreted was published in a top medical journal.”

Meanwhile, the CDC website portrays the Danish RCT (with its 4,800 participants) as being far less relevant or important than the observational study of Missouri hairdressers with no control group, dismissing the former as “inconclusive” and “too small” while praising the latter, amazingly, as “showing that wearing a mask prevented the spread of infection”—when it showed nothing of the sort.

Each of the RCTs discussed so far, 13 in all, examined the effectiveness of *surgical* masks, finding little to no evidence of their effectiveness and some evidence that they might actually increase viral transmission. None of these 13 RCTs examined the effectiveness of *cloth* masks. “Cloth face coverings,” according to former CDC director Robert Redfield, “are one of the most powerful weapons we have.”

One RCT tested these masks that so many high-profile public-health officials have touted. This “first RCT of cloth masks,” in the trial’s own words (it is apparently still the only one), was a 2015 [study](#) by MacIntyre, et al. in Hanoi, Vietnam. A relatively large study, with over 1,100 participants, it tested cloth masks against surgical masks and did not feature a no-mask control group. The trial tested the protection of health-care workers, instructing them to wear a two-layer cloth mask at all times on every shift (“except in the toilet or during tea or lunch breaks”) across four weeks.

The study found that those in the cloth-mask group were 13 times more likely (2.28 percent to 0.17 percent) to develop an influenza-like illness than those in the surgical-mask group—a statistically significant difference. The trial also lab-tested penetration rates and found that while surgical masks were “poor” at preventing the penetration of particles—letting 44 percent through—cloth masks were “extremely poor,” letting 97 percent through. (N95 hospital respirators let 0.1 percent through.)

The authors write that wearing a cloth mask “may potentially increase the infection risk” for health-care workers. “The virus may survive on the surface of the facemasks,” they explain, while “a contaminated cloth mask may transfer pathogen from the mask to the bare hands of the wearer,” which could lead to hand hygiene being “compromised.” As for double-masking, the authors write, “Observations during SARS suggested double-masking . . . increased the risk of infection because of moisture, liquid diffusion and pathogen retention.” Absent further research, they conclude, “cloth masks should not be recommended.”

MacIntyre and several other authors of this study, perhaps under pressure from the CDC or other entities with similar agendas, released what the CDC calls a “follow up study,” in September 2020. This follow-up isn’t really a study at all, certainly not a new RCT, yet the CDC cites it favorably while disparaging the original study, which, the CDC asserts, “had a number of limitations.” This 2020 follow-up pretty much amounts to publishing the finding that when hospitals washed the cloth masks, health-care workers were only about half as likely to get infected as when they washed the cloth masks themselves. Still, the 2020 publication says, “We do not recommend cloth masks for health workers,” much as the 2015 one said.

Other reviews of the evidence have been mixed but generally have come to similar conclusions. Certain masking advocates [admit](#) that the RCT evidence is “inconclusive” but cite other forms of evidence that have held up [poorly](#). A [study for Cochrane Reviews](#) by Jefferson, et al. that examines 13 of the 14 RCTs discussed herein (all but the Denmark Covid-19 study) notes “uncertainty about the effects of face masks” and writes that “the pooled results of randomised trials did not show a clear reduction in respiratory viral infection with the use of medical/surgical masks during seasonal influenza.” Meantime, [a study by Perski, et al.](#), which performed a Bayesian analysis on 11 of the 14 RCTs discussed herein, concluded that when it comes to “the benefits or harms of wearing face masks . . . the scientific evidence should be considered equivocal.” They write, “Available evidence from RCTs is equivocal as to whether or not wearing face masks in community settings results in a reduction in clinically- or laboratory-confirmed viral respiratory infections.”

In sum, of the 14 RCTs that have tested the effectiveness of masks in preventing the transmission of respiratory viruses, three suggest, but do not provide any statistically significant evidence in intention-to-treat analysis, that masks might be useful. The other eleven suggest that masks are either useless—whether compared with no masks or because they appear not to add to good hand hygiene alone—or actually counterproductive. Of the three studies that provided statistically significant evidence in intention-to-treat analysis that was not contradicted within the same

study, one found that the combination of surgical masks and hand hygiene was less effective than hand hygiene alone, one found that the combination of surgical masks and hand hygiene was less effective than nothing, and one found that cloth masks were less effective than surgical masks.

Hiram Powers, the nineteenth-century neoclassical sculptor, keenly observed, “The eye is the window to the soul, the mouth the door. The intellect, the will, are seen in the eye; the emotions, sensibilities, and affections, in the mouth.” The best available scientific evidence suggests that the American people, credulously **trusting** their public-health officials, have been blocking the door to the soul without blocking the transmission of the novel coronavirus.

Jeffrey H. Anderson served as director of the Bureau of Justice Statistics from 2017 to 2021, and is co-creator of the Anderson & Hester Rankings, part of college football's Bowl Championship Series formula from 1998 to 2014.

Photo by Michael M. Santiago/Getty Images

City Journal is a publication of the Manhattan Institute for Policy Research (MI), a leading free-market think tank. Are you interested in supporting the magazine? As a 501(c)(3) nonprofit, donations in support of MI and *City Journal* are fully tax-deductible as provided by law (EIN #13-2912529).

DONATE

52 Vanderbilt Avenue New York, NY 10017 | (212) 599-7000