STAT

Do masks work? Randomized controlled trials are the worst way to answer the question

By Baruch Fischhoff, Martin Cetron *and* Katelyn JetelinaMay 2, 2023

Mark Lennihan/AP

Early in the Covid-19 pandemic, before we had vaccines and effective medical procedures, the only ways to prevent transmission of the virus were behavioral measures like face masks and social distancing. There

was (and continues to be) a desperate hunger for definitive studies telling us how well specific measures would work, with specific publics, in specific settings, for specific strains of the novel and changing virus. Although the scientific community mobilized at record speeds, it could not produce studies with the desired surety.

Indeed, the resulting studies could not, on their own, produce the evidence that public policy demanded. The virus was too new, and there were too many questions to answer all at once. As a result, unless researchers took extraordinary care to ensure that their necessarily ambiguous results were not misinterpreted, these studies led to incorrect conclusions — especially in a politicized environment.

Such research sows confusion that erodes trust in science, misleads policymakers, depletes social capital, and squanders critical resources. We believe that many of these studies should never have been done at all, reserving resources for studies that could improve health outcomes. Comprehensive reviews of these studies show how little many could teach us.

As an example, a review published in the Lancet examined hundreds of observational studies of the health, economic, and educational effects of medical and behavioral interventions, with an emphasis on their distributional (also known as social justice) impacts. The review summarized an extraordinary evidentiary base, reflecting the scientific skills and public resources invested in the research and the review. It found suggestive evidence that jurisdictions that were sensitive to historical health inequities achieved better health outcomes.

However, as the authors of the review acknowledge, the results from observational studies are inherently ambiguous because they are correlational. For example, some U.S. states fared much better than others. Those states were also more likely to implement some policies,

like mandatory masking in public, than others. However, they were also different in many other ways, such as their socioeconomic resources, population dispersion, public health infrastructure, political alignment, and leadership. The authors cite these confounds when tempering their social justice policy recommendations. Critics cite the confounds when savaging the recommendations.

One way to avoid these confounds could be a randomized controlled trial (RCT) that divides people into otherwise equivalent groups that either do or do not receive an intended treatment. Even more robust evidence comes from meta-analyses, which are statistical summaries that combine the results from all available RCTs. Meta-analyses give greater weight to the better studies, such as ones with consistently administered protocols and larger samples. But even these sometimes come with major limitations.

Cochrane has published [two major meta-analyses of RCTs](#) evaluating field trials of face masks involving the general public. Both meta-analyses have been widely misinterpreted as showing that face masks don't work. What they really show is that the RCTs asked questions that they could not [answer](#). Cochrane's leadership recognized these limits in an editorial accompanying the first meta-analysis, recommending that policymakers rely on other sources of evidence.

The researchers conducting these RCTs doubtless designed the best studies that they could, given the field circumstances, available resources, and investigator capabilities. But what if their best wasn't good enough? What if it is so difficult to conduct scientifically sound randomized trials of mask wearing that even the best studies reveal little? Such studies can confuse people who want to know how effective face masks are, while emboldening people who are already completely convinced that face masks are ineffective — and are looking for grounds to sow doubt about them.

According to a recent article from the Gates Foundation, "As of 2022, at least 329,830 trials were in progress, seeking to learn whether a new drug, vaccine, device, procedure, or behavior is better than the alternatives." These trials involve legions of researchers and millions of research subjects. Proper randomization reduces the chance that groups differ for reasons such as genetics, environment, and health equity.

However, there's an enormous problem lurking here. The designs of most clinical trials are too weak to answer the question that they pose — namely, whether an intervention succeeded. The Gates Foundation article notes that only an estimated 5% of RCTs for Covid-19 drugs were designed to yield statistically meaningful results. Such "uninformative research" wastes precious time and money, while incurring the incalculable opportunity costs of diverting resources from research that might advance public health and increasing the risk of chance "positive" results that lead down dead ends.

RCTs have value only when researchers can be sure that the treatment is administered as intended. With an RCT for a drug, that means knowing, for example, whether providers' biases affected who got the drug, whether patients' habits affected how they took it, and whether control group participants somehow got it on their own. Without that knowledge, an RCT produces noise, and meta-analyses produce piles of noise.

With behavioral interventions like wearing masks, it may be impossible to produce anything *but* noise without vastly more ambitious studies than have been conducted to date.

Three years ago, we knew very little about how and how well face masks work against this disease. Today, we have strong evidence regarding the effectiveness of face masks in the form of laboratory

studies, theoretical analyses, and RCTs that involved health care personnel. It has not come from RCTs of face masks distributed to the general public. However, the good evidence from other sources has been largely drowned in the noise created by flawed reports of RCTs involving the general public, studies that could not possibly have sent a clear signal.

Thus, one strong lesson from the current pandemic is to invest in carefully designed field trials of behavioral interventions that, combined with other evidence, can produce clear enough policy signals to move the field forward, especially during a crisis. That requires significant coordination of the scientific field and a realistic assessment of the investigators' ability to design and implement the intervention faithfully and monitor adherence, in both the treatment and the control group.

Following that research policy will likely mean concentrating scientific resources in a very small number of excellent, likely multi-center, studies that serve the diverse interests of different publics and policymakers, with different concerns for disease, economics, education, and equity. It might even mean conducting no field trials at all, unless researchers can demonstrate their ability to help, rather than harm. There is no point in trying to evaluate the effectiveness of face masks directly, unless researchers can demonstrate who wore masks, when, and how well, for both the treatment and control group.

Instead of fatally flawed RCTs, we may need to rely on two sources identified in the November 2020 Cochrane editorial. One source is field trials in which behavior can be observed and controlled. Such trials have found that requiring high-quality, well-fitted masks in hospitals reduces disease transmission. That evidence gives reason to hope that face masks will benefit ordinary people wearing imperfect, imperfectly fitted masks, under everyday circumstances.

The second alternative source is studies of factors that might affect mask efficacy. For example, how well do various kinds of mask block virus-sized particles in laboratory tests that simulate inhaling and exhaling? How well can people put on various masks, with various kinds of instruction? When do people wear masks in various real-life settings (stores, restaurants, buses, planes, airport terminals)? Are people wearing them to protect themselves or others? How is mask wearing affected by what other people are saying, and doing?

Moreover, members of the public need communications that help them follow the science that is worth following. That means responsible reporting, not hyping individual studies that did not, and could not, tell us anything.

Scientists need self-discipline to avoid conducting studies that cannot answer the question that they ask. They can only exercise that discipline if academic institutions reward the "slow science" of such research, focused on the public good rather than institutional budgets and professional resumes. Uninformative research causes significant harms: eroding trust in science, misguiding policymakers, depleting social capital, and squandering critical resources — all leading to poorer health outcomes.

Baruch Fischhoff is a decision scientist at Carnegie Mellon University and member of the National Academy of Sciences and National Academy of Medicine. Martin Cetron is an infectious disease epidemiologist and former director of global migration and quarantine at the Centers for Disease Control and Prevention. Katelyn Jetelina is an epidemiology, data scientist, and science communicator who publishes Your Local Epidemiologist.

About the Authors

Baruch Fischhoff

Do masks work? RCTs can't tell us - STAT

https://www.statnews.com/2023/05/02/do-masks-work-rcts-randomiz...

baruch@cmu.edu

Martin Cetron

linkedin.com/in/marty-cetron-0008925/

Katelyn Jetelina

2 Comments

Create a display name to comment

This name will appear with your comment