It’s a wash: Hands-on hygiene in Peru

By Kai Kupferschmidt

Few public health interventions promise an easier payoff than getting people to wash their hands with soap. It’s easy, and studies suggest it could prevent up to two-fifths of all diarrheal disease and one-fourth of pneumonia, the two biggest childhood killers in poor countries, saving up to 1 million lives every year. With hand-washing, there’s none of the controversy surrounding condoms and birth control. And rigorous attempts to measure impact can cause unease among major donors, the groups they fund to roll out programs, and disease advocates, says Ruth Levine, a development economist at the William and Flora Hewlett Foundation in Menlo Park, California, who edited the first edition of Millions Saved. “The support for global health rests on a collective hope that money is turning into lives saved, and anything that punctures that belief is really very threatening,” Levine says.

Until 2000, hardly any impact evaluations were done in global health, or for that matter in development aid in general. “If you asked anyone what their impact was—and I don’t care whether it was diabetes, hypertension, HIV—the answer would have been, ‘We’re spending X amount of dollars,’” says Mark Dybul, who heads the Global Fund to Fight AIDS, Tuberculosis and Malaria, which was formed in 2002.

The Global Fund and the U.S. President’s Emergency Plan for AIDS Relief (PEPFAR), which started in 2003 and was later headed by Dybul, together have spent more than $60 billion on HIV/AIDS, and both have received flak for not taking a close enough look at their own impact. For instance, they have long used the number of patients given antiretroviral drugs as a major yardstick of success. But people don’t always take their pills, or they may drop out of treatment. So the precise public health impact, as well as the cost-effectiveness of specific programs, remained unclear.

Even if public health does improve after the rollout of a program, there may be no causal relationship. What’s needed for a thorough evaluation, says epidemiologist Nancy Padian, who has appointments at both the Berkeley and San Francisco campuses of the University of California (UC), is a way to assess what would have happened if the intervention had not occurred. This is known in the lingo of impact evaluations as a counterfactual, and it’s akin to a placebo control in a drug trial. “It’s all about having the most robust counterfactual you can have,” says Padian, who was a lead scientific adviser for PEPFAR.

A landmark demonstration of the value of this approach involved a social welfare program launched in Mexico in 1997, called PROGRESA, in which families received cash
for keeping their kids in school and using preventive health services. PROGRESA’s main architect, Mexican Deputy Finance Minister Santiago Levy, was worried that the next government might shut the initiative down unless hard evidence showed that it worked to improve kids’ health. Levy enlisted an evaluation team led by UC Berkeley health economist Paul Gertler.

Gertler proposed taking advantage of the fact that Mexico could not afford to roll out PROGRESA nationwide all at once. He suggested a lottery to determine which communities could participate in the program first. Other villages would start 2 years later; they became the counterfactual. The comparison showed that the cash transfer led to a significant drop in illness and hospital visits among children, and adults benefited, too. The study was “a phenomenal breakthrough,” Levine says, and PROGRESA survived.

RANDOMIZED, CONTROLLED STUDIES create their own counterfactual by randomly assigning participants to intervention or control groups; a mainstay of clinical research, they have rapidly multiplied in global health (see graphic). One leader is the Abdul Latif Jameel Poverty Action Lab (J-PAL), founded in 2003 at the Massachusetts Institute of Technology to do such studies in health and other development projects. J-PAL has since done follow-up evaluations of PROGRESA (renamed Oportunidades) and studied the impacts of hand-washing promotion on diarrhea in Peru, double-fortified salt on anemia in India, and deworming on school attendance in Kenya. 3ie, founded 5 years later, was established to fund evaluations and serve as an online repository of high-quality studies.

While randomized, controlled trials may be the ideal for measuring impact, they aren’t always feasible. It’s also widely considered unethical to withhold a proven intervention—some of them life-saving—from one group of people simply to test how well a large-scale rollout works. So scientists have developed several less rigorous methods. Eligibility criteria, for example, have a built-in counterfactual: If an intervention applies to people only under 14, kids who just turned 15 become good comparators. Sophisticated techniques can also “match” the group receiving an intervention to an artificial counterfactual created by statisticians. “There is a suite of methods,” Padian says, some of which aren’t used in standard drug or vaccine trials but are considered convincing to many, including the CGD editors of Millions Saved.

The Global Fund and PEPFAR have embraced the idea of more rigorously measuring impact. “It’s becoming a top priority,” says Deborah Birx, who heads PEPFAR. Today, both programs want to track how many people on HIV treatment have fully suppressed the virus for prolonged periods, which means they’re actually healthier. To accomplish this, PEPFAR is distributing computer tablets to clinics and asking them to record—and report in real time—HIV levels in patients on treatment. Dybul says he’s also “really adamant right now” about finding out how specific clinics are doing, instead of focusing on national data. That’s a great step forward, says Stefano Bertozzi, dean of the School of Public Health at UC Berkeley.

“If you know you have clinics in one country that go from 25% to 90% of patients being virally suppressed, as a manager, you have incredible information to know what’s working and what isn’t,” Bertozzi says.

But the rising popularity of impact evaluations has triggered plenty of debates, which some have slagged as “wont work.” A single impact study often doesn’t mean much because the results may not apply elsewhere, says Harvard University economist Lant Pritchett, a prominent critic of impact evaluations and a nonresident fellow at CGD. The randomistas tend to overlook the “key failing” in developing countries, he says: Organizations don’t