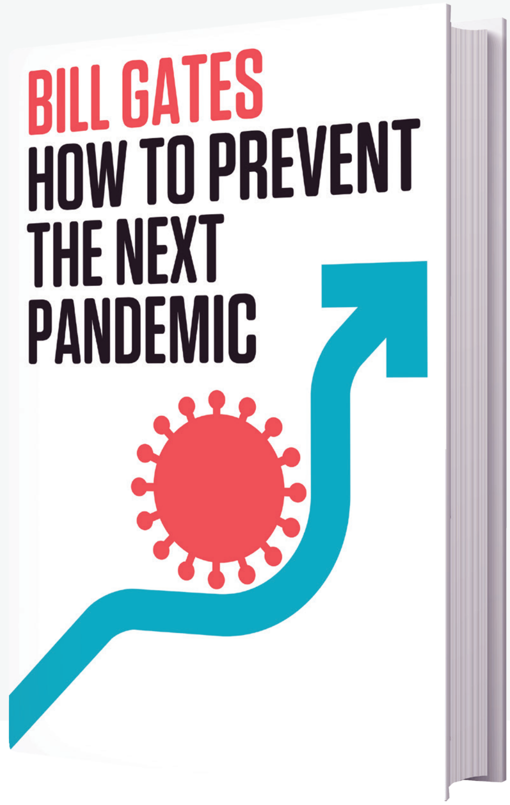
Gates Notes

A CHAPTER FROM

How to Prevent the Next Pandemic

Available to purchase
U.S. Retailers
International Retailers



CHAPTER 7
Practice, practice, practice

In July 2015, *The New Yorker* published an article that got attention up and down the West Coast of the United States. I live just outside of Seattle, and I remember emailing the article to friends just as it was arriving in my inbox from other friends. It became a regular source of dinner-table conversation that summer.

The headline on the article was "The Really Big One: An earthquake will destroy a sizable portion of the coastal Northwest. The question is when." The author, a journalist named Kathryn Schulz who won the Pulitzer Prize for this piece, explained that a huge stretch of coastline, from Canada into Washington state, Oregon, and northern California, sits near what's known as the Cascadia Subduction Zone. Cascadia is a fault hundreds of miles long beneath the Pacific, where two tectonic plates meet and one is sliding underneath the other.

Subduction zones are inherently unstable and tend to cause earthquakes. Seismologists figure that massive earthquakes occur along the Cascadia zone an average of every 243 years, and that the last one occurred around 1700. The 243-year average is debated, and Cascadia may go much longer than that between quakes, but when we read the article, none of us locals could dismiss the fact that the last Cascadia quake happened more than 315 years ago.

The article cited horrific projections: A Cascadia quake, and the tsunami that would result from it, could kill nearly 13,000 people, injure 27,000 more, and displace a million people from their homes. And the toll could be much worse if the quake occurred during the summer, when tourists crowd the beaches of the West Coast.

To test how ready the Pacific Northwest is for the really big one, the federal government oversees a series of periodic full-scale exercises known as Cascadia Rising. The 2016 exercise involved thousands of people from dozens of government agencies, the military, nonprofits, and businesses. A lengthy after-action report detailed the results and issued a series of lessons learned during this real-life exercise. Among other things, the report noted, "Catastrophic response requirements are fundamentally different than any response we have seen before. . . . A massive response will be required." Another Cascadia Rising exercise is scheduled for summer 2022.

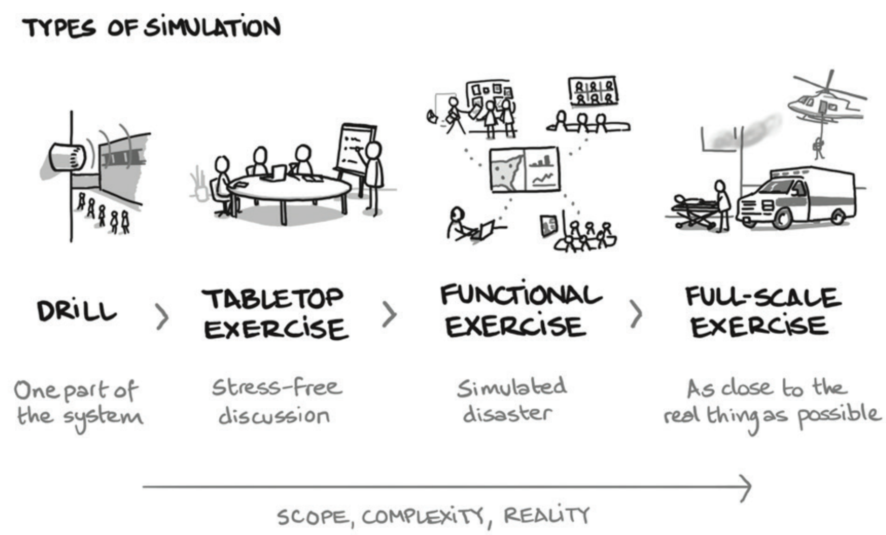
I wish I could report that Cascadia Rising has led to major changes and the Pacific Northwest is now as ready as it can be for a catastrophic quake. Unfortunately, that's not the case. For one thing, retrofitting all or even most of the region's buildings to make them seismically sound would be prohibitively expensive.

But the exercises are still worthwhile. At least the government is trying to get people to focus on the problem.

We tend to use words like *drill* and *exercise* interchangeably, but in the world of disaster preparation, they don't mean the same thing at all.

A drill is a test of just one part of a system—say, whether the fire alarm in your building works, and whether everyone knows how to get out quickly.

Then, moving up the scale of complexity, there's the tabletop exercise, a discussion designed to identify and solve problems. More complicated still is the functional exercise, a simulated disaster that tests how well the whole system operates, but without moving people or equipment.



Finally, there are full-scale exercises, such as Cascadia Rising. These are designed to get as close to the real thing as possible— complete with actors pretending to be sick or injured and vehicles moving people and equipment around.

For as long as I've been learning about pandemic preparedness and prevention, I've been amazed that there isn't an ongoing series of full-scale exercises designed to test the world's ability to detect and respond to an outbreak. As the WHO's flu preparedness program put it in a 2018 guide to running outbreak exercises, "Considerable efforts and resources have been invested by countries around the world in developing national pandemic influenza preparedness plans and the capacities needed to respond to an influenza pandemic. However, to be effective, plans need to be tested, validated and updated periodically through simulation exercises."

There have been many tabletop and functional exercises for disease outbreaks, but perhaps only a handful of country-scale ones designed to simulate an outbreak of flu or coronavirus.* Credit for running the first one seems to belong to Indonesia, which held a full-scale outbreak exercise in Bali in 2008. There have been no exercises involving entire regions throughout the world.

Although details are sometimes murky because governments make some of the results classified—especially for full-scale exercises—it appears that the history of these simulations is spotty. On the positive side is Vietnam, which has held frequent simulations at various levels of complexity, taken action to fix problems that were revealed, and set itself up to respond especially well to COVID.

But often, in other countries, these exercises end with a series of what-ifs and missed opportunities.

For example, the United Kingdom ran an exercise called Winter Willow in 2007 and another, Cygnus, in 2016, both focused on flu outbreaks. Cygnus in particular highlighted problems with the government's readiness and produced a series of classified recommendations that went unheeded—and which caused a scandal when *The Guardian* revealed them during the first year of the COVID pandemic.

The United States had a similar experience in 2019, when the government ran Crimson Contagion, a series of exercises designed to answer one question: Was the country ready to respond to an outbreak of a novel flu virus?

Overseen by the Department of Health and Human Services, Crimson Contagion occurred in two phases. The first involved a series of seminars and tabletop exercises held between January and May, in which people from all levels of government, plus the private sector and nongovernmental organizations, got together to discuss existing plans for responding to an outbreak.

In the second phase, they put these plans to the test in a functional exercise. Over four days in August 2019, participants worked through a scenario in which tourists visiting China have become ill with a respiratory illness caused by a virus. They fly out of the airport in Lhasa and travel to other cities in China before flying home to their respective countries.

The virus turns out to be just as contagious as the 1918 flu strain, and only slightly less deadly. It rapidly spreads from one human to another, making its first U.S. appearance in Chicago and quickly moving to other major cities.

At the beginning of the exercise, it has been forty-seven days since the first U.S. case. There are moderate or high caseloads across the Southwest, Midwest, and Northeast. Models are predicting that the virus will make 110 million people in the U.S. sick, put more than 7 million in the hospital, and kill 586,000 Americans.

Over the next four days, the participants would debate decisions that would have been unfamiliar to anyone not already versed in outbreak response work: quarantines, personal protective gear, social distancing measures, school closures, public communications, the purchase and distribution of vaccines. Today, of course, these terms are part of our everyday vocabulary.

The scope of Crimson Contagion's functional exercise was enormous. It involved 19 federal departments and agencies, 12 states, 15 tribal nations and pueblos, 74 local health departments, 87 hospitals, and more than 100 groups from the private sector. When it was all over, the participants got together to discuss how it had gone. While

**Exercises related to animal-borne diseases aren't unheard of. For example, four years after a disastrous outbreak of foot-and-mouth disease in 2001, the United Kingdom and five Nordic countries ran simulations to test their readiness.*

they found a few things that worked well, they found many more that did not. I'll mention just a few, which will sound eerily familiar.

No one in the exercise understood what the federal government was responsible for, versus what others would do. The Department of Health and Human Services didn't have clear authority to lead the federal response. There wasn't enough money to buy vaccines (in this scenario, there was already an available vaccine for the strain in question, but it hadn't been administered). State leaders didn't know where to turn for accurate information. There were huge discrepancies in how states planned to deploy scarce resources, such as ventilators, and some had no plan at all.

Some of the troubles were almost comically mundane, like something out of the TV show *Veep*. Federal agencies confused participants by unpredictably changing the names of conference calls. Sometimes the name of the meeting used some unrecognizable acronym, so people wouldn't show up. State governments, already understaffed, struggled to keep up with all the calls while also handling the response itself.

It is telling that in the official government report on the results of Crimson Contagion—dated January 2020, just as COVID cases were starting to mount—the word *diagnostics* appears only three times over fifty-nine pages. The report simply notes that diagnostics will be one of many supplies that will be hard to come by in a pandemic. Just a few weeks later, of course, the United States' inability to ramp up testing in a serious way would become tragically apparent. It bears repeating: America's failure to test people at anything near the level that other nations achieved is among the biggest mistakes any country has made during the pandemic.

Crimson Contagion was not the first simulation designed to test America's readiness to handle an outbreak. That honor likely goes to a tabletop exercise with the ominous name Dark Winter, which was held over the course of two days in June 2001 at Andrews Air Force Base in Washington, D.C.

Surprisingly, Dark Winter wasn't organized by the federal government, but by independent organizations whose leaders were growing increasingly concerned about the potential of a bioterror attack on the United States and wanted to draw attention to the problem.

Dark Winter supposed that a terrorist group releases smallpox in Philadelphia, Oklahoma City, and Atlanta, infecting a total of 3,000 people. Less than two months later, the disease had spread to 3 million people and killed one million, with no end in sight. An observer I know commented that the outcome was Smallpox 1, Humanity 0.

Other exercises followed: Atlantic Storm in 2005 (another smallpox attack), Clade X in 2018 (an outbreak of a novel influenza virus), Event 201 in 2019 (an outbreak of a novel coronavirus), a simulation at the Munich Security Conference in 2020 (a bio-attack involving an engineered influenza virus).*

Even though each of these U.S. exercises imagined different scenarios, were run in different ways, and used different methods, they had three things in common. One is that their conclusions are fundamentally the same—there are huge gaps in America's and much of the world's ability to contain outbreaks and prevent a pandemic— and they propose various ways to fill those gaps.

The second thing these exercises have in common is that none of them led to any significant changes that

**The Gates Foundation was one of the funders of the Event 201 exercise. Some conspiracy theorists suggested that it predicted COVID. As the organizers made clear, it wasn't a prediction, and they said so at the time. You can find a statement about this at centerforhealthsecurity.org.*

would make America better prepared for an outbreak. Although some adjustments were made at the federal and state levels, we only need to look at what's happened since December 2019 to see that whatever changed was not enough.

The third is that, with the exception of Crimson Contagion, each of the U.S. simulations took place exclusively in conference rooms, and none of them involved moving real people or equipment from one place to another.

Full-scale exercises aren't run as often as tabletop and functional ones for the obvious reason that they're expensive, time-consuming, and intrusive. In addition, some public health leaders have argued that the best way to prepare for a pandemic is to simulate smaller outbreaks, which means not preparing for things that happen only in an epidemic or pandemic—problems like supply chains being disrupted, economies shutting down, and heads of state interfering for political reasons. It's also likely that, until 2020, the threat of a worldwide contagion seemed remote to most people, and therefore not worth the trouble and cost of a full-scale, real-life exercise.

Two years into COVID, the argument is much easier to make. The world needs to be running far more full-scale exercises that test its readiness for the next major outbreak.

In most countries, these exercises can be run by national public health institutions, emergency operations centers, and military leaders, with the GERM team that I described in Chapter 2 acting as an advisor and reviewer. For low-income countries, the world will have to bring in resources to help out.

Here's how a full-scale outbreak exercise might work. The organizers would pick a city and act as if it's experiencing a bad outbreak that could spread nationally or globally. How quickly can a diagnostic test for the pathogen be developed, manufactured at a high volume, and delivered wherever it is needed? How well, and how quickly, can the government get accurate information out to the public? How do local health officials handle quarantines? And—as we now know can happen—what if supply chains are cut, local health agencies make poor decisions, and political leaders interfere?

They would establish a system for reporting cases and running genetic sequencing on the pathogen. They would recruit volunteers to try out nonpharmaceutical measures, modify them based on how the disease spreads, and understand the economic impact they would have during a real emergency.

And if the pathogen initially spreads through human contact with animals, the exercise would evaluate a government's ability to dispose of the animals.* Suppose it's an avian flu spread by chickens: Because so many people rely on chickens for their livelihood, they'll be reluctant to slaughter the birds on the off chance they might spread a flu. Does the government have the money to compensate them for their losses, and a system for doing so?**

To make the exercise even more realistic, software would generate surprise events from time to time, throwing a wrench into the plan to see how everyone responds. Software would also be used to track the overall

**In November 2020, the Danish government ordered the culling of 15 million mink out of concern about a COVID mutation that might move from them to humans.*

***If you want more detail from experts on what an exercise might involve, see the WHO document "A Practical Guide for Developing and Conducting Simulation Exercises to Test and Validate Pandemic Influenza Preparedness Plans," available at who.int.*

simulation and to record actions for later review.

In addition to advising countries on their simulation plans, the GERM team would measure readiness in other ways—for example, by looking at how well a given country's health system is detecting and responding to nonpandemic diseases. If it's a place where malaria is a problem, how early does the system detect big outbreaks? Or with tuberculosis and sexually transmitted diseases, how well does it trace the recent contacts of people who test positive? On their own, these proxies wouldn't tell researchers everything they need to know, but they would offer clues about weaknesses in the system that need more attention. The countries that do a good job of watching for, reporting, and managing endemic diseases are in a good position to respond to a pandemic threat.

The GERM team's most important role will be to distill the findings from exercises and other measures of preparedness, record the recommendations that come out of them—ways to strengthen supply chains, better methods for coordinating across governments, agreements to improve the distribution of medicines and other supplies—and then try to keep the pressure on world leaders to translate these findings into action. We've already seen how little things changed after Dark Winter, Crimson Contagion, and the other outbreak simulations. Unfortunately, there's no innovation that can make sure that after-action reports don't simply get stored on some website and then forgotten. Political leaders and policymakers will need to change this.

To get a sense of the different scales at which full-scale exercises can be run, let's look at two examples from disaster preparedness, starting with a relatively small one.

In the summer of 2013, Orlando International Airport in Florida simulated a horrific aviation-related disaster, an exercise designed to meet the federal government's requirement that all U.S. airports run a full-scale simulation once every three years. In the scenario, according to an article in *Airport Improvement* magazine, a hypothetical jetliner carrying ninety-eight passengers and crew members experiences hydraulic problems and crashes into a hotel a mile from the airport.

The exercise involved 600 volunteers pretending to be victims, 400 first responders, and staff from sixteen hospitals, and it took place in a training facility with three aircraft and a four-story building designed to let firefighters practice on real fire. Officials had to establish who was in command. First responders had to triage patients, treat the ones they could, and transport others to the hospital. Security had to manage a crowd of observers. Friends and family of the victims had to be notified. News reporters needed updates. The exercise identified some necessary improvements and cost about $100,000.

At the other end of the complexity spectrum, there's the full-scale exercise run by U.S. military forces in August 2021. Over the course of two weeks, personnel from the Navy and Marine Corps participated in the largest naval training event in a generation. The name of this exercise—Large-Scale Exercise 2021—understated its scope. Simulating concurrent wars with two world powers, LSE 2021 spanned seventeen time zones and involved more than 25,000 personnel, using virtual reality to allow participants to join remotely and to link units from around the world so they could share information in real time.

The analogy between war games and germ games is not perfect. Stopping an outbreak, after all, is different from fighting a war. Countries should be working together, not against one another. And unlike military exercises, outbreak simulations can involve the public and be highly visible so they're no more out of the ordinary than a fire drill.

Still, the ambition of the LSE is impressive. The exercise created the opportunity for organizations spread around the world to share data and make fast, informed decisions together. It is hard to read about that and not think: *We need something like this for pandemic prevention*.

A good model of a simulation is a full-scale exercise developed by Vietnam in August 2018, designed to see how well the system identified a potentially worrisome pathogen. I'm impressed by how meticulous it was.

Four actors were hired to play patients, family members, and their contacts, and they were given scripts with key information for the medical staff (who knew they were participating in an exercise). On Day 1, the actor portraying a fifty-four-year-old businessman arrived at the emergency room of a hospital in the northeastern province of Quang Ninh, complaining of a dry cough, fatigue, muscle pain, and shortness of breath. The doctor questioned him thoroughly enough to discover that he had recently traveled to the Middle East, where he could have picked up the MERS virus—a fact that, in combination with his symptoms, was enough to get him admitted to the hospital and isolated.

News of the worrisome case made its way up the chain of command within minutes, and soon the members of a rapid response team had arrived at the hospital and at the man's residence. The actors were tested using throat swabs, which were then replaced with samples that had been spiked with the virus that causes MERS. Although the samples weren't actually driven to a lab, the organizers waited the length of time it would have taken to transport them before lab staff ran real tests and correctly identified the positive MERS cases.

The exercise didn't go off flawlessly—the organizers noticed a number of gaps in the process—but it would be surprising if it had been flawless. The point is that the gaps were identified and, most important, fixed.

This full-scale exercise was small, compared with the national and regional ones that the world needs, but it had many of the necessary components. If exercises like this one were run by more countries and in more regions, they would keep us from making a classic mistake: preparing for the last war.

It will be tempting to assume that the next major pathogen will be as transmissible and as lethal as COVID, and as susceptible to innovations like mRNA vaccines. But what if it isn't? There is no biological reason why the next pathogen couldn't be far more lethal. It could quietly infect millions of people before a single person starts feeling sick. Our bodies might not be able to knock it out with neutralizing antibodies. With germ games, we'll be able to test against the wide range of pathogens and scenarios that the next outbreak might present.

Since the risk of a pandemic is higher than the risk of an all-out war, we should be running an LSE-sized global exercise organized by the GERM team at least once a decade. Each region should run another major exercise in the same decade, with advice from GERM, and countries should undertake smaller simulations with their neighbors.

There is one reason to hope that the reports generated by future exercises won't be ignored: experience. In the early days of COVID, many experts thought the countries that had gone through the SARS outbreak in 2003 were better prepared for this pandemic. Having experienced how bad it was, the theory went, they were ready—politically, socially, and psychologically—to do what it took to protect themselves. The theory proved true. The places hit hardest in 2003 included mainland China, Hong Kong, Taiwan, Canada, Singapore, Vietnam, and Thailand. When COVID emerged, most of these places responded quickly and decisively, limiting the numbers of COVID cases for more than a year.

Maybe Crimson Contagion, Dark Winter, and the rest didn't have more impact because their scenarios seemed so remote at the time—at least to most people and most politicians. Now, though, the idea of a virus spreading around the world, killing millions of people and doing trillions of dollars in damage, is very real for all of us. We should take the outbreak of a disease at least as seriously as we take earthquakes and tsunamis. To keep a pandemic like COVID from happening again, we need to practice stopping pathogens early, learn which parts of the system need to improve, and be willing to change even when it's difficult to do so.

So far in this book, I've stuck to writing about naturally occurring pathogens. But there is another, even more unsettling scenario that disease exercises must account for—a pathogen that's intentionally deployed with the goal of killing or maiming huge numbers of people. In other words, bioterrorism.

The history of turning viruses and bacteria into weapons stretches back centuries. In 1155, Frederick I, the Holy Roman Emperor, laid siege to the town of Tortona (in modern-day Italy) and is said to have poisoned the local water wells with dead human bodies. More recently, in the eighteenth century, British soldiers distributed blankets used by smallpox patients to Native Americans. In the 1990s, members of the Aum Shinrikyo cult released sarin gas in the Tokyo subway, killing thirteen people, and reportedly released botulinum toxin and anthrax four times without causing any casualties. And in 2001, a series of attacks using anthrax sent through the U.S. mail left five people dead.

Today the natural pathogen that would make the most fearsome weapon is surely smallpox. It's the only human disease that has ever been eradicated from the wild, though samples are still kept in government labs in the United States and Russia (and possibly in other countries too).

What makes smallpox especially scary is that it spreads fast through the air and has an extremely high mortality rate, killing around a third of everyone who's infected. And because most vaccination programs stopped after it was eradicated in 1980, almost no one is immune to it any longer. The United States does have a stockpile of smallpox vaccines large enough to protect everyone in the country, but as we've seen with COVID vaccines, distributing the doses would not be a simple matter—especially when people are panicking about an attack—and it's unclear how the rest of the world could be protected.

Part of the risk stems from the fall of the Soviet Union. As my friend Nathan Myhrvold notes in his paper "Strategic Terrorism," an international treaty banned bioweapons in 1975, but the Soviet Union continued its program into the 1990s—"thereby producing thousands of tons of weaponized anthrax, smallpox, and far more exotic biological weapons based on genetically engineered viruses."

The chances that terrorists will get their hands on these existing weapons are compounded by the fact that the science behind engineering pathogens is no longer the sole province of highly trained scientists working in secretive government programs. Thanks to the advances in molecular biology of the past few decades, students at hundreds of colleges and universities around the world can learn everything they need to know to engineer a biological weapon. And some scientific journals have published information that a terrorist could use to concoct a new pathogen, a practice that has led to vigorous debate about how to share research knowledge without adding to the risk.

We haven't yet seen a mass attack using an engineered bioweapon, but it is certainly not out of the question. In fact, during the Cold War, Soviet and American labs produced bioengineered anthrax that was resistant to antibiotics and evaded every vaccine. A nation-state or even a small terrorist group that developed smallpox resistant to treatment and vaccines would be capable of killing over a billion people.

A new pathogen could be designed that is highly communicable and lethal but doesn't cause symptoms right away. Such a pathogen would spread quietly around the world, perhaps for years, before arousing suspicion. HIV, which evolved naturally, works this way; although people can infect others very quickly after acquiring it, their health may not fail for nearly a decade, allowing the virus to go undetected while they spend years passing it to others. A pathogen that operated this way but didn't require intimate contact to spread, as HIV does, would be far worse than the AIDS pandemic.

"To put it in perspective," Nathan writes, a single attack that caused 100,000 casualties "would kill more people than were killed cumulatively in all terrorist actions by all parties throughout history. It might take anywhere from 1,000 to 10,000 typical suicide bombings to equal it." It is this scale of catastrophe—the kinds of events that can kill hundreds of thousands, millions, or even billions of people—that deserves far more attention than it gets.

Now, I am an optimistic person who's naturally inclined to focus on solutions. But even I have to admit that it's hard to write a list of responses that feels adequate to the threat of bioterrorism. Unlike a natural pathogen, an intentionally made disease can be designed to get around our tools of prevention.

All the things we need to do to prepare for a deliberate attack are a super-set of those we need to do to prepare for a natural one. Outbreak exercises can focus on attack scenarios and test our readiness. Better treatments and vaccines are important, no matter what the source of the pathogen is. Better diagnostics that return a result in thirty seconds would make it more practical to screen people at airports or public events, where we'd be most likely to see the spread of an engineered pathogen, and of course they would be extremely useful for everyday testing as well. Mass-scale genomic sequencing of pathogens will help in an ordinary flu outbreak and during an attack. Even if an attack never comes, we'll be glad to have all these tools available.

We also need some approaches that are specifically designed to deal with deliberate attacks. I am hopeful that we'll have devices in airports and other big gathering places that detect pathogens in the air and sewage, but the technology is still years away. The U.S. government made an attempt at a much larger-scale version of this approach in 2003 with a program called BioWatch, which placed devices designed to detect airborne anthrax, smallpox, and other pathogens in cities throughout the country.

Although BioWatch still operates in twenty-two states, it is widely regarded as a flop. Among other faults, it relies on the wind blowing in exactly the right direction and takes up to thirty-six hours to confirm a pathogen. Sometimes the detectors fail to work for the most basic reason: They get unplugged.

Regardless of whether air-sniffing machines have a future, the possibility of a bioterror attack is another reason why the world should be putting far more money and effort into research on detecting, treating, and preventing diseases that can go global. Given the national security implications of an attack and the chance that the number of casualties could reach into the millions, more of this research should come from defense budgets. The Pentagon's budget is roughly $700 billion a year, while the National Institutes of Health budget is about $43 billion a year. As far as resources are concerned, the Department of Defense operates on a whole other level.

Although I'm optimistic that science will deliver better tools for stopping outbreaks from any source, governments should also consider a defense that is as low-tech as they come: a reward. There's precedent for it—governments frequently offer to pay people for information leading to the arrest of criminals and terrorists. Considering the scale of damage that can be done these days, governments should be willing to pay quite a lot to informants who help thwart a bio-attack.

Regardless of what the final bioterror plan looks like, it will need to survive shifting political winds. In the early 1980s, while he was running the CDC, Bill Foege worked with the FBI on a program for detecting and responding to bioterrorism. The program included simulations of attacks using different diseases, to see how such attacks would work, as well as a defensive plan for each disease. Foege's successor, convinced that such an attack would never happen, closed down the program. If the U.S. and the rest of the world make a big investment in germ games and get the public's attention, it will be much harder for a single political appointee to get in the way of protecting people.