To be Precise
One Friday afternoon, Andrea Baccarelli, MD, PhD, and his research team got to talking about what types of publicly available data could explain the geographic disparities in extreme aging documented by the World Health Organization and others.

It was 4 p.m. when they commenced digging around online. They started by sifting through Census Bureau reports, looking for communities where disproportionate numbers of people live long enough to celebrate their 85th birthday, the benchmark for extreme aging. Then they pulled up other data—NASA satellite maps showing global air pollution, U.S. Centers for Disease Control and Prevention analyses of obesity prevalence, and Institute for Health Metrics and Evaluation reports on cigarette smoking. By Monday at noon, they had the framework for a report printed by *Environmental Health Perspectives* in November 2016. County by county, they found that local pollution levels were more predictive of longevity than smoking, obesity, or socio-demographic features like race or income.

Americans spend more of our gross domestic product on health than any nation in the world. Yet we suffer shorter lifespans and poorer quality of life. In its Healthy People 2020 guide to reversing those trends, CDC details 1,271 objectives spanning 42 topic areas. Setting community-level public health policy in response is like trying to take a sip from a firehose. Investigators must build the evidence base to help individuals and communities narrow their focus, says Baccarelli, chair and the Leon Hess Professor of Environmental Health Sciences. “We can create tools,” he says, “to help communities decide on top priorities and calculate the top interventions.”

That, in a nutshell, is the definition of precision public health—providing the right intervention to the right population at the right time. Focusing on the relevant data can save governments money and redirect scarce public dollars to where they’re needed most. Public health scholars have a surfeit of raw material with which to work, and Mailman School investigators are on it, devising a wealth of strategies to achieve the promise of precision prevention.
Scientists have long known that air pollution and other chemical exposures wreak havoc on the human body. Trained as a clinical endocrinologist, Baccarelli—who heads the Mailman School’s Laboratory of Environmental Precision Biosciences—has dedicated his career to understanding how. His early research sought clues to the process by which chemicals interact with genes to disrupt the endocrine system. Over the last decade, Baccarelli’s focus has narrowed in on epigenetics—the biochemical pathways that translate genetic codes.

Baccarelli was among the first to show that pollutants alter epigenetic translation through a process called DNA methylation, in which a molecule known as a methyl group attaches itself to DNA, silencing some genes and activating others. Since 2010, he has published more than 170 papers, many of which have been highly cited, exploring all manner of environmental exposures: air pollution, metals, bisphenol A and other chemicals, and psychosocial stress.

His work has also laid the groundwork for public health policy. Baccarelli’s 2008 report showing that prenatal dioxin exposure affects thyroid function in children was one of two analyses considered by the Environmental Protection Agency when it lowered the legal limit for dioxin emissions in the U.S. More recently, his documentation of the health effects of pollution in Beijing has informed the Chinese government’s increasingly strict clean-air regulations.

While the epigenome is often considered most vulnerable during the prenatal period, pollutants inflict damage at any age. In a study of air pollution exposure in older men, Baccarelli was among the first to show that our epigenome changes as we age. “There is evidence,” he says, “that older individuals are much more vulnerable to air pollution.”

Meanwhile, Baccarelli is exploring avenues for prevention and treatment. A pair of papers published this spring in *PNAS* and *Science Reports* details a clinical trial testing whether vitamin B supplements can reverse epigenetic damage caused by air pollution. Baccarelli and his coauthors found that the vitamins nearly erased the cardiovascular and immune damage caused by fine-particle air pollution. “Our study launches a line of research for developing preventive interventions to minimize the adverse effects of air pollution,” he says. “Because of the central role of epigenetic modifications in mediating environmental effects, our findings could very possibly be extended to other toxicants and environmental diseases.”
The Virus Hunters

W. Ian Lipkin, MD, director of the Mailman School’s Center for Infection and Immunity (CII) and the John Snow Professor of Epidemiology, builds new tools to make disease detection faster and simpler in a bid to better contain epidemics.

This summer, the Federal Drug Administration issued an emergency use authorization for a test developed by a team led by CII associate research scientist Nischay Mishra, PhD, that can simultaneously detect the presence of Zika virus and all serotypes of dengue virus, chikungunya virus, and West Nile virus in up to 88 samples of blood in less than two hours. The test can also be used to detect Zika in urine. Known as the ArboViroPlex test, for the four types of arbovirus it detects, the assay builds on nearly two decades of research by Lipkin and his colleagues to speed the diagnosis of pathogens early in an outbreak, when it’s easiest to curb.

In 2003, Lipkin played a pivotal role in curbing the SARS virus by creating a test—using a technique called polymerase chain reaction—that amplifies viral DNA for detection even in blood samples with minimal infection. Because his test could also detect the product of this amplification as it was happening, it could churn out diagnoses far more quickly than other methods.

In the intervening years, Lipkin has improved upon the technology he developed to detect SARS to create a test that can screen a single sample for the presence of hundreds of viruses that affect humans and animals. Known as VirCapSeq-vert, the tool contains 2 million genetic probes that detect the signature DNA of all known vertebrate viruses. Each probe binds to a corresponding viral sequence; when a particular virus is present in a sample, a magnetic process “pulls out” its unique sequences, which can then be used to identify the virus.

The platform could revolutionize how first responders react to deadly outbreaks like Ebola, dengue fever, or even the flu by accelerating diagnosis. If you add the three together, you can capture the entire universe of infections,” Lipkin says. “So it’s really quite powerful.”
Power is exactly what investigators seek when they sift through big data. The more data points researchers can consider, the more likely they are to find patterns, trends, and associations that may elucidate disease mechanisms. And new technology can capture all kinds of information—medical histories, geographic locations, biometric data, lifestyle habits—that can then be combined with millions of other bytes of information. It’s no easy task, though, to figure out how to meaningfully analyze such gargantuan data sets. Biostatisticians approach the problem by turning the data inside out and upside down. They make tables, plot graphs, account for outliers and missing data points, and write and rewrite algorithms to sort the array of information that streams from smartphones and brain scans.

F. DuBois Bowman, PhD, the Cynthia and Robert Citrone-Roslyn and Leslie Goldstein Professor and chair of Biostatistics, creates, refines, and employs such quantitative tools to discover disease biomarkers in huge data sets. His current research focuses on making sense of the millions of measurements of brain activity that can be collected through such high-tech imaging technologies as functional magnetic resonance imaging (fMRI), diffusion tensor imaging, and positron emission tomography.

His analysis of brain scans of people with and without Parkinson’s disease, for instance, experimented with methods to cull relevant insights from thousands of data points for each research participant. Bowman had just four participants, but their fMRI records—continual imaging that reveals the moment-to-moment changes in neural activity that occur even when a person is resting—generated 46,580 measures for him to assess. So he designed an algorithm to detect correlations—first among brain regions, then among subregions, and finally among voxels (the fundamental, three-dimensional units within an fMRI image, which each contain just a million or so brain cells).

In a 2016 study published in *Frontiers in Neuroscience*, Bowman and his coauthors detailed the 24 structural and functional differences between the healthy participants and those with early stage Parkinson’s. Talk about a needle in a haystack: Those 24 relevant structural and functional differences represented just 0.051 percent of the original data.

“I like to say the brain boggles the mind,” Bowman says. “It’s very complicated. With our analytic methods, we try to characterize properties that are associated with brain function, but it is a daunting problem. Research in how to do this well continues to grow and evolve.”
And so, too, does the volume of information available. With the worldwide explosion of mobile technology such as smartphones, researchers can now tap into a continuous data stream about all kinds of human behaviors—our movements around town, our shopping habits, even the influence of our social networks on our mood and health behaviors.

That continuously growing store of data demands new methods of real-time analysis and optimization, the focus of work by Professor of Biostatistics Ken Cheung, PhD. An expert in artificial intelligence and machine learning, Cheung has devised analytical methods to test and enhance the personal relevance of IntelliCare, a suite of smartphone apps designed to promote mental health.

Each app is based on a clinically proven intervention for anxiety and depression. Using artificial intelligence algorithms, the software learns from how users interact with a given app, then recommends other apps within the suite that users might also find helpful. Cheung aims to enhance the software’s responsiveness by analyzing thousands of seconds of user-interaction data and rewriting the underlying code to leverage the accumulated knowledge of an entire population of users to personalize care.

Those personalized recommendations are designed to maximize user engagement, says Cheung, because increased user engagement with the effective mental health practices IntelliCare supports has been shown to correlate with improvements among people who suffer from anxiety and depression. Cheung has already written a preliminary algorithm. Compared with the original software’s results, his version boosted the number of meaningful user interactions by ten sessions per week. “Eventually,” he says, “the goal is to have the recommendations informed by user data in such a way that they update in real time. So the recommendations will get better and better.”

The benefit of Cheung’s work isn’t limited to IntelliCare and its suite of apps, says the professor. The research and algorithms he and his team have developed can also be applied more broadly, creating new opportunities to mine big data for precise, effective interventions to inform and improve individual lives.
In 1994, the state of Florida launched an experiment to overhaul welfare. The five-year randomized trial put time limits on recipients’ eligibility for benefits and offered job training and other employment-related support. Two decades later, on a quest to document the long-term health effects of that reform, Peter Muennig, MD, MPH ’98, professor of Health Policy and Management, compared the state’s data on the people involved in the experiment with its 2011 Social Security rolls. What he and his colleagues found was alarming: In a 2013 paper published in Health Affairs, they reported that hundreds of former welfare recipients had died prematurely.

The investigators uncovered the deadly side effect by zeroing in on the post-welfare circumstances of women who were unable to join the workforce due to disability or family-care responsibilities. Compared with those who found paid work when their benefits ran out, the women who couldn’t work were at greater risk of homelessness and more likely to die sooner. “Failing to target programs to specific groups can be harmful,” says Muennig. “In the case of welfare reform, the ‘average’ person benefited, but some people died because the program didn’t consider that anyone would be harmed. That’s bad policy.”

For more than two decades, Muennig—who is also a member of the Mailman School’s Global Research Analytics for Population Health program—has sifted through the aftermath of policy change to uncover the nuances lawmakers might miss, quantifying both return on investment and such unintended consequences as the premature deaths of the women in Florida.

In ongoing work, he continues to explore the health outcomes of state-level welfare reforms. He’s also investigating whether government-funded prekindergarten programs designed to boost cognitive development among low-income children also improve their health. The latter draws upon decades of follow-up data on kids who attended programs for particularly impoverished families in the 1960s and 1970s. Muennig found that as adults they had more-stable family environments, better health insurance coverage, higher earnings, and healthier behaviors.

So when is pre-K a smart public health investment? Muennig is currently analyzing additional data collected by the U.S. Department of Health and Human Services. “It looks like if you target these programs to the kids who are really disadvantaged, you get these big benefits,” he says. “But if you don’t, you don’t.”
In 1975, famed Columbia epidemiologists Mervyn Susser and Zena Stein led a study of IQ among survivors of the Dutch Hunger Winter, a six-month famine that began in September 1944 when Nazi troops blocked food supplies to the Netherlands’ western provinces. Women who became pregnant during that brutal winter subsisted on as little as 500 calories per day. While Susser and Stein found no intelligence deficits among the tens of thousands of adults who were either conceived or in the early stages of fetal development during the Hunger Winter, they did find a higher rate of neural tube defects such as spina bifida and microcephaly.

More than four decades later, their son—Ezra Susser, MD, MPH ’82, DrPH ’92, professor of Epidemiology and of Psychiatry—has found epigenetic clues from the same cohort that suggest concrete maternal health interventions to promote improved psychiatric health among the next generation. The work began with a study of psychiatric hospitalization records, which revealed an increased risk of schizophrenia among adults conceived just before or during the famine. Subsequent studies pointed to a role for the IGF2 gene, whose expression is modified within the cohort.

More recently, analyses of the Autism Birth Cohort (ABC) Study—a trove of data compiled from thousands of Norwegian children with autism and their mothers—homed in on folate’s role in the prenatal pathways of cognitive development. A micronutrient found in dark green leafy vegetables, folate protects against neural tube defects during the first trimester—precisely the diagnoses the elder Susser and Stein found in their Dutch Hunger Winter analyses. In the Norwegian study, children of women who had insufficient pre- and early-pregnancy folate levels had a greater chance of being diagnosed with autism. Ezra Susser wants to know whether a mother’s folate levels might also be related to her children’s chances of developing schizophrenia. The answer won’t come quickly. The ABC babies are still decades away from developing symptoms, which rarely manifest in childhood.

Meanwhile, Susser and his colleagues are examining the DNA in cord blood samples collected from some of the ABC children, looking for epigenetic tags similar to those seen in the Hunger Winter cohort. “It will take some time to figure this all out,” says Susser. “I hope not my whole lifetime.”

NANCY AVERETT writes about science and the environment from Cincinnati. Her work has been published in Pacific Standard, Audubon, Discover, and others.