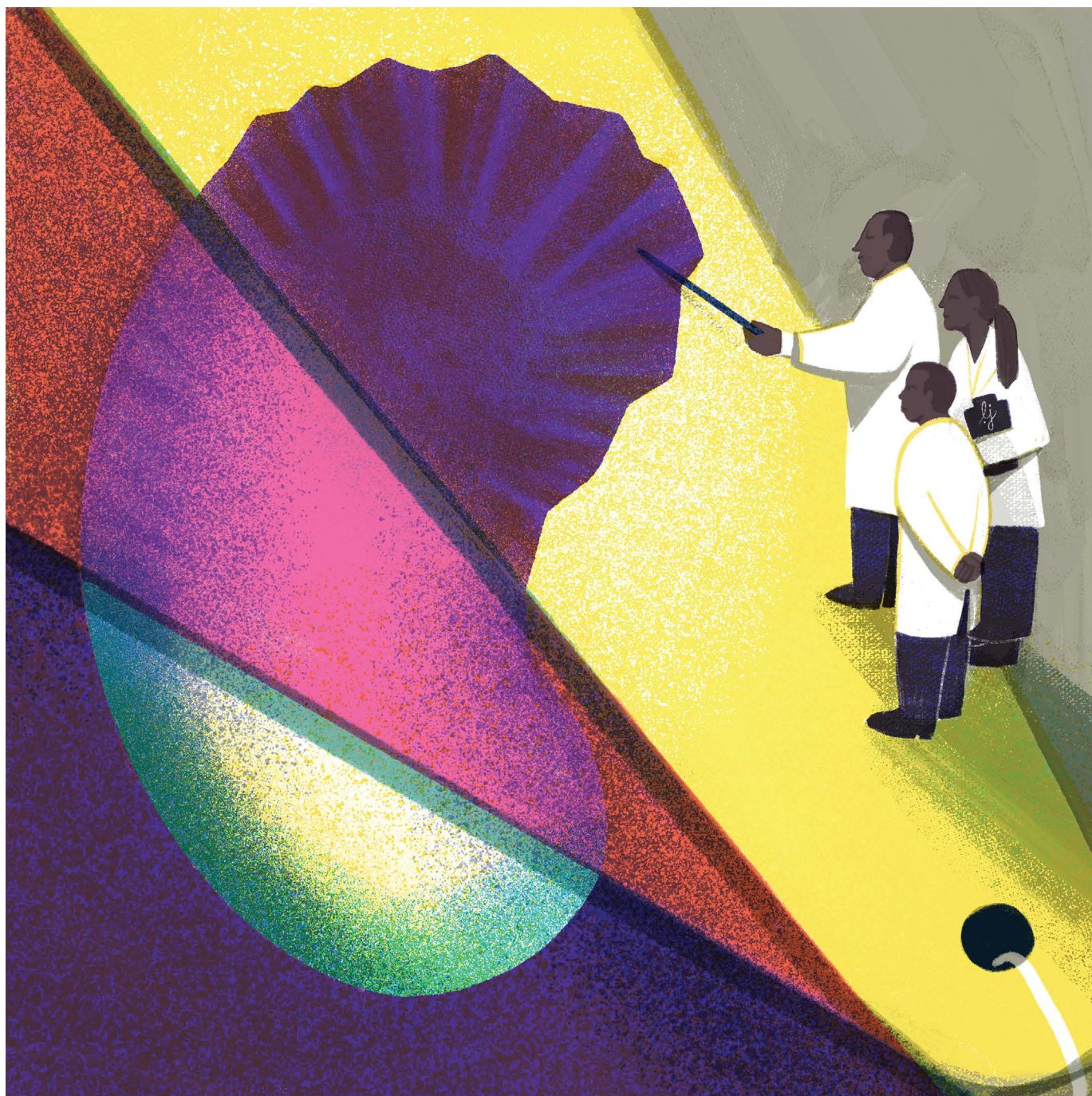




Better Measures

New devices, formulas and tools are removing historical bias from medical diagnoses BY CASSANDRA WILLYARD

MELANIE HOENIG WAS TEACHING first-year medical students how to estimate kidney function when one of them, Cameron Nutt, raised his hand. Why, he asked, did the diagnostic algorithm include an adjustment for Black patients? In the U.S., Black people have higher rates of kidney disease and kidney failure and are less likely to get

a kidney transplant than white people, but the adjustment makes it seem as though Black people have *better* kidney function than people of other races who have the same test results.

Good question, thought Hoenig, a kidney specialist at Beth Israel Deaconess Medical Center in Boston. She had never

wondered why this might be. “I said, ‘You’re right. That doesn’t make any sense,’” Hoenig recalls of the 2016 classroom conversation.

This value for kidney function, called the estimated glomerular filtration rate (eGFR), helps doctors figure out when to send patients to a specialist, when to start dialysis, when they are eligible to join the wait list for a kidney transplant, and where their name lands on that list. Adjusting the algorithm for Black patients decreased their chances for treatment and transplant.

The equations and instruments doctors

rely on are infused with historical bias. Medicine has long treated race as though it provides important information about the underlying biology and genetics of disease, a strategy that has had an enormous impact on diagnosis and treatments. People have been passed over for kidney transplants, denied therapies and diagnosed with diseases later than necessary simply because of the color of their skin.

Race is a social construct that reveals little about ancestry. There is more genetic variation within racial groups than between them. “The racial differences found in large datasets most likely often reflect effects of racism—that is, the experience of being Black in America rather than being Black itself,” researchers wrote in a 2020 *New England Journal of Medicine* article outlining the dangers of race-adjusted algorithms.

To undo this bias, researchers are changing the algorithms and instruments and finding new models to reduce disparities.

KIDNEYS FILTER WASTE and excess water from the blood through tiny structures called glomeruli. Directly measuring how well these glomeruli are functioning is possible but cumbersome, so instead doctors rely on blood levels of a protein called creatinine, a waste product produced by muscles and a by-product of protein metabolism, to estimate the glomerular filtration rate (GFR). When kidneys are working well, they filter out creatinine; if the kidneys start to fail, creatinine levels rise. The protein is easy and inexpensive for laboratories to measure.

The first equation to assess kidney function, developed in the 1970s, relied on age, sex, weight and creatinine levels in the blood. But the formula wasn’t precise. So, in the late 1990s, a team of researchers set out to develop a more accurate one. They used existing data from a study of creatinine and GFR in more than 1,600 people, then correlated the two measurements. The team looked at 16 different factors that might influence the relationship. (We tend to lose muscle mass as we age, for example, so older people have lower creatinine levels than younger people.) The authors noted that for any given GFR, creatinine was higher in Black people than in white people. Why that might be wasn’t clear. Maybe it was because Black people had higher muscle mass, they

speculated. The study population was only 12 percent Black, yet the difference felt too substantial to ignore.

To account for this difference, the researchers added an adjustment for Black patients: a multiplication factor of up to 1.21, which essentially inflated their estimated kidney function by as much as 21 percent. In 2009 the researchers published an updated equation, but the Black correction factor remained, albeit lower, up to 1.16. “We always recognized that race was not the biological process by which African Americans differed from non-African Americans in the relationship between GFR and creatinine,” Andrew Levey, who worked to develop both equations, later explained. But “it stood in for something that was important.”

“The way the lab report was written was, if your creatinine is a 4.0, your kidney function is 19 percent. Oh, unless you’re African American; then it’s 22 percent,” says Martha Pavlakis, a nephrologist at Beth Israel Deaconess. “It makes no sense.” In people with healthy kidneys, small differences don’t matter. But when kidney function declines, eGFR, which decreases as blood creatinine levels rise, becomes crucial. That number helps to determine whether a patient is referred to a nephrologist, diagnosed with kidney disease or deemed eligible to join the wait list for a kidney transplant.

Hoenig began working with a small group of students from Harvard Medical School’s Racial Justice Coalition to lobby to eliminate the correction factor, and in 2017 Beth Israel Deaconess became the first medical center to do so. Efforts elsewhere largely stalled until the deaths of George Floyd, Ahmaud Arbery and Breonna Taylor, three Black Americans whose deaths made national news. In the wake of their killings, conversations about race rippled throughout the medical community, Pavlakis says.

As protests erupted across the country, medical students and faculty at many major universities began to circulate petitions calling for an end to the use of the racial correction in eGFR. Some major academic health systems began removing race from the equation, but their approaches were inconsistent. Neil Powe, chief of medicine at Zuckerberg San Francisco General Hospital and Trauma Center, and other experts watched the changes unfold with concern. There was no

unified way of diagnosing kidney disease. “You could be at one hospital and have a diagnosis of kidney disease. You go down the street [to another hospital], and you wouldn’t have kidney disease,” Powe says. “That was just chaos.”

In the summer of 2020 the National Kidney Foundation and the American Society of Nephrology formed a task force to assess how best to move forward. “They thought we’d solve it overnight, but it took us about 10 to 11 months to churn through this,” says Powe, who co-led the task force. Ultimately they chose an equation that used the same 2009 data but eliminated race as a variable, then refit the curve to the whole dataset.

A conversation about race was also happening at the Organ Procurement and Transplantation Network (OPTN), which manages transplants from deceased donors. The wait list for a kidney is long. Patients aren’t eligible to join until they meet certain criteria; these can vary at different transplant centers, but all candidates must have an eGFR of 20 percent or less. And because of the eGFR correction factor, Black patients needed higher creatinine levels than people of other races to pass that threshold. “Nobody who came up with the formula was like, let’s keep Black people off the list. But that, in fact, was the result,” Pavlakis says.

In July 2022 the race variable was explicitly forbidden in organ allocation. Pavlakis saw that as just the first step. She wanted to help Black patients already on the list and those who had previously been denied entry because of their kidney function numbers.

In January 2023 the OPTN decided that transplant centers should look back at the lab reports of Black patients on the list and recalculate their eGFR using the race-neutral equation to see whether they should have been referred for transplant. “Basically, half the Black patients on the transplant list got extra priority added to their standing because of this project,” Pavlakis says.

Pavlakis acknowledges that this change doesn’t fix every disparity in kidney allocation. But she also sees it as restorative justice. “It’s not perfect,” she says, but “I think it’s probably the largest example of fixing a race disparity that is out there.”

PULMONOLOGISTS have been grappling with a similar problem. To assess lung function, doctors ask patients to blow into a de-

vice called a spirometer, which measures the maximum amount of air a person can exhale and how much they can force out of their lungs in a single second. The spirometer compares those numbers with reference values for “normal” lung function. The results help doctors diagnose diseases such as emphysema and chronic obstructive pulmonary disease, assess severity of those conditions and monitor declines in lung function.

What constitutes “normal” varies by age, sex, height and, until recently, race. Why race? Data collected in the late 1800s and early 1900s suggested different races have different lung capacities, a phenomenon researchers ascribed to innate biology rather than social, economic or environmental factors. By the early 20th century the idea that lung capacity varied among racial groups was “an ostensible fact,” wrote Brown University researcher Lundy Braun in a 2015 article on the historical use of race in spirometry. What experts missed was that race was probably a proxy for other factors, such as air quality, nutrition, and other exposures, that affect lung health and development.

When the European Respiratory Society’s Global Lung Function Initiative developed reference values for spirometry in 2012, it used more than 160,000 spirometry results from 33 countries. Researchers observed “proportional differences in pulmonary function between ethnic groups” and decided to develop separate values for four groups: Caucasian, African American, North Asian and Southeast Asian. They also used an “other” category for people who didn’t fit elsewhere. The model assumes that, compared with white adults, Black adults have about 10 to 15 percent smaller lung capacity and that adults of Asian ancestry have 4 to 6 percent smaller lung capacity. So the same spirometry results in Black, Asian and white people led to different interpretations of health. As a result, lung diseases in certain populations have gone undiagnosed and untreated.

The division of reference values by race is problematic for many reasons. “We’re a big melting pot,” says Alexander Niven, a pulmonologist at the Mayo Clinic in Minnesota. So even if there were “a specific cluster of genes that predispose people to greater or less lung function, that’s highly unlikely to remain a pure cluster in this global world.”

What’s more, lungs are in constant con-

“Half the Black patients on the transplant list got extra priority added to their standing because of this project.”

—Martha Pavlakis *Beth Israel Deaconess Medical Center*

tact with the outside world and continue developing throughout childhood and into early adulthood, Niven says. “It’s impossible to separate race from all of these other factors that unfortunately are inexplicably linked to different populations within our society, many of which are likely coloring the changes in lung function that we see in different social groups.”

In practice, the race-based model doesn’t seem to improve predictions when it comes to outcomes that matter. “You can’t tell any better who’s going to go to the hospital. You can’t tell any better who’s going to die. You can’t tell any better who has severe symptoms and who doesn’t. And in some of those cases, you actually worsen your ability to predict by adding race,” says Aaron Baugh, a pulmonary and critical care physician at the University of California, San Francisco.

In 2023 the Global Lung Function Initiative replaced race-based equations with a race-neutral equation. That same year the American Thoracic Society and the European Respiratory Society recommended all health-care providers switch to the new formula.

That shift is happening now, and researchers are just beginning to uncover the broad impact of this change. “Long story short, it’s profound,” says Arjun Manrai, a bioinformatics researcher at Harvard Medical School. Lung function helps to determine disability payments, candidacy for some professions, priority for lung transplants, and more. Manrai and his colleagues found that some 10 million people in the U.S. would have their diagnosis or the severity of their disease reclassified. Disability payments could increase by more than \$1 billion. Such changes are not always beneficial. A new diagnosis can make someone ineligible for certain jobs, such as firefighting. And a Black person with lung cancer might not be identified as a good candidate for surgery because their lung function may be too poor to allow for removal of part of their lung. “There are trade-offs essentially attached to these reclassifications,” Manrai says.

The new equation comes from the same 2012 data as the original formula, and it isn’t perfect. “We kind of settled on the race-neutral equations we have now as the best current option, knowing that in the future, something better might arise,” Baugh says.

MANRAI THINKS A LOT about how traditional algorithms operationalize race, adjusting what constitutes “normal” for any particular patient, and how lessons from those algorithms can be incorporated into producing more sophisticated machine-learning algorithms. “They can be biased, and they can propagate the very same sort of race-based medicine,” he says. “But they’re a tool, and the tool can also be used in the reverse direction: to mitigate existing disparities and to potentially reduce existing biases in the health-care system.”

One example of how AI might help improve health equity is evident in research on disparities in knee pain. Previous studies have shown that Black people routinely report more intense knee pain from arthritis than people of other races. But often that pain can’t be explained by the structural damage visible in x-rays. As a result, it is often dismissed or attributed to external factors such as psychological stress.

Emma Pierson, who studies machine learning and health-care inequities at Cornell University, and her colleagues wanted to understand whether there might be physical signs in the knee itself that could explain this pain disparity. They used knee radiographs and patient pain scores from more than 4,000 people who had osteoarthritis or were at risk of developing it to train a machine-learning model.

Surprisingly, the model predicted pain better than the traditional arthritis scoring system. Specifically, Pierson says, “it seems to be picking up on factors that disproportionately affect underserved patients.” What those factors might be isn’t clear, and Pierson emphasizes a need for caution. “In general, the capabilities of these models tend to out-

strip our ability to understand how they're achieving those capabilities," she says.

SOMETIMES DIAGNOSTIC INSTRUMENTS introduce bias. The fingertip clamps doctors use to measure oxygen levels in the blood, for example, work by measuring the absorption of different wavelengths of light to estimate the blood oxygen level. But the device, called a pulse oximeter, tends to overestimate oxygen saturation in people with darker skin tones.

Researchers have known about this problem for decades, but manufacturers didn't feel much pressure to fix the problem. The effect was relatively minor, and it was most prominent at low oxygen saturations. "That difference was probably correctly assumed to not be physiologically relevant," says Michael Lipnick, an anesthesiologist at the University of California, San Francisco, who leads a research project to assess pulse oximeter performance. "If somebody's oxygen saturation is really 1 percent or even 2 percent higher or lower than the real value, there's no harm."

When the COVID pandemic sickened millions of people, however, small biases had an outsize effect. "Clinical decisions were being made based on that number," Lipnick says. In 2023 a team of researchers looked at health records from more than 24,000 people hospitalized with COVID during the first 19 months of the pandemic. They zeroed in on those who had both a pulse oximeter reading and an arterial blood gas test, the gold standard for measuring oxygen saturation in the blood. Pulse oximeter readings consistently overestimated oxygen levels in Black and Hispanic patients. Black patients were also more likely than white patients to have their need for COVID therapy underestimated because of inaccurate pulse oximeter readings. Such oversight has clinical consequences: being passed over for COVID treatment resulted in an hour's delay in care on average and a higher risk of readmission.

Lipnick is part of the Open Oximetry Project, which has been testing different pulse oximeters in diverse groups to get a sense of their real-world performance. He and his colleagues have seen a range of variability. Most devices tended to perform worse when used on people with darker skin pigment, but some performed better.

Researchers are working to develop more accurate tools, and regulators are considering larger test populations with a variety of skin tones. Lipnick wants better pulse oximeters but worries that some of the fixes may increase costs. "It's a big concern, especially in low- and middle-income countries, where the majority of the world's people with darker skin pigment live," he says.

In the short term, Lipnick says, clinicians should rethink how they use data from pulse oximeters. "It gives a number, and we assume that that number is truth." In reality, the number might be off by as much as 5 percent. If doctors recognize the error rate, they can make decisions that aim to minimize health-care disparities. "I think a lot of the solution will lie

in how we use the technology," he says.

Pavakis also sees a need for more critical thinking on the part of clinicians. She is dismayed at the number of years that she relied on the eGFR equation without stopping to carefully consider the rationale for its race correction. "When we were taught this formula, we were like, 'This is data-driven. This is from a research study. This must be accurate,'" she says. Evidence-based, however, doesn't always mean equitable, and that's the real goal. Hoenig's students and other people who recognized bias are making health care better for all.

Cassandra Willyard is a science journalist based in Madison, Wis. She covers public health, medicine, and more.

Rural Prescriptions

Some of the most innovative solutions for health care come from rural regions around the world BY CARRIE ARNOLD

ON A FRIGID WINTER EVENING about five years ago, a desperately ill young woman walked through the doors of the Sanford Bemidji Medical Center in rural Minnesota. Several weeks before, she had labored alone for hours in her tiny mobile home to bring a new baby into the world. The woman had received no prenatal care and no medical attention at delivery—the kind of situation that has made maternal mortality rates for Native American women in rural areas twice as high as those of white women. The only reason she was showing up now was that the baby wasn't eating. She had no running water to make formula. The hospital was her only option. Johnna Nynas, the obstetrician on call, quickly diagnosed her patient with postpartum preeclampsia, a rare condition that affects people after pregnancy and can be deadly if untreated.

For Nynas's pregnant patients, the hospital in Bemidji is the only option between Duluth, Minn. (three hours away), and Fargo, N.D. (2.5 hours away). The surrounding area is one of the poorest in Minnesota. Some residents of the nearby Leech Lake, Red Lake and White Earth Indian Reservations don't have reliable access to running water. With so many pressing un-

met needs, people find it difficult to get prenatal care. Transportation (especially in winter) and child care for medical visits that require a several-hour car ride and possibly an overnight hotel stay are often unaffordable, even if Medicaid covers the cost of the health care. Nynas, who was born and raised in rural Minnesota, says that by the time an expectant parent arrives in her office, they may have a list of health concerns that have gone untreated for years. She links this lack of care directly to the elevated risk of pregnancy-related deaths and complications in the region.

"When we first meet patients, it's probably the first contact they've had with the health-care system in quite some time," Nynas says. Haunted by her patient's preeclampsia emergency, she set out to remove barriers to needed care. Loaned blood pressure cuffs and bathroom scales let many of her low-risk patients receive checkups over the phone. This communication made it easier to schedule in-person visits for ultrasounds and blood tests.

David Driscoll, director of the Healthy Appalachia Institute at the University of Virginia, isn't surprised that the impetus for change began in a rural area. The regions